THE UNIVERSITY of EDINBURGH

| | |
|---|---|
| **Title** | Quantum brane cosmology |
| **Author** | Seery, David |
| **Qualification** | PhD |
| **Year** | 2004 |

# Quantum brane cosmology

David Seery

Submitted for the degree of Doctor of Philosophy

The University of Edinburgh, 2004

# Contents

Declaration 1

Acknowledgements 3

Introduction 5

**Part 1. Background: Cosmology and Quantum Mechanics** 13

Chapter 1. Low energy physics: notation and conventions 15

1.1. Einstein gravity 16

1.1.1. Coupling to matter 17

1.2. Spinors 18

1.3. The $(n+2)$ description of curvature 21

Chapter 2. Quantum field theory 23

2.1. Classical mechanics and classical field theory 24

2.2. Quantum mechanics 26

2.3. Quantum field theory 30

2.3.1. Path integral transition amplitudes 31

2.3.2. The expansion into diagrams 32

2.3.3. Perturbative renormalization 37

2.3.4. The quantum effective potential 45

2.4. Quantization of gauge field theories 47

2.4.1. Constrained Hamiltonian systems and reduced phase space 49

2.4.2. The Fadeev–Popov determinant 52

2.5. BRST symmetry 56

Chapter 3. String theory, compactification, and membranes 61

3.1. The Polyakov action and the string spectrum 64

3.1.1. The string spectrum 67
3.1.2. Strings in background fields 74
3.2. Compactification to low dimensions 75
3.2.1. The wave operator and the Kaluza–Klein mechanism 76
3.3. T-duality and strings at strong coupling 78
3.3.1. Closed strings 78
3.3.2. Open strings 81
3.3.3. Strings at strong coupling 82
3.4. M-theory and the Hořava–Witten theory 84

Chapter 4. Cosmology 87
4.1. Introduction 87
4.2. Homogeneous and isotropic cosmologies 90
4.2.1. The Friedmann equation 90
4.2.2. Thermodynamics in an expanding universe 92
4.2.3. Decoupling and freeze-out 97
4.2.4. Recombination 99
4.2.5. Primordial nucleosynthesis 100
4.3. Large scale structure formation 102
4.3.1. The density perturbation 103
4.3.2. Perturbation theory 105
4.3.3. The cold dark matter transfer function 110
4.3.4. The variance $\sigma_R$ 112
4.4. The cosmic microwave background 112
4.4.1. The Sachs–Wolfe integral 113
4.5. Inflation 116
4.5.1. Introduction 116
4.5.2. Scalar field cosmology 120
4.5.3. Inflation and transplanckian physics 123
4.6. Perturbations and the origin of structure 124
4.6.1. The gauge-invariant description of perturbations 125
4.6.2. Perturbations from scalar field inflation 128
4.7. The no-hair theorem 143

4.8. Dark energy and the anthropic landscape 145
4.9. Observational summary 151
4.9.1. Galaxy surveys 153
4.9.2. Cosmic microwave background 154

Chapter 5. Brane cosmology 157
5.1. Binétruy–Deffayet–Ellwanger–Langlois models 160
5.1.1. The background BDEL metric 161
5.1.2. The metric fields $a$ and $n$ 164
5.2. Heterotic M-theory and moving brane models 167
5.2.1. The effective five-dimensional action 169
5.2.2. Cosmological solutions and moving branes 173
5.3. Ekpyrotic and cyclic models 175
5.4. Verlinde compactification 179
5.4.1. The hierarchy problem 182
5.5. The brane Einstein equations 183
5.5.1. Cosmological solutions 188
5.6. Stability of the brane 189
5.6.1. Tachyon matter and brane instabilities 191
5.7. BDEL brane compactifications and zero modes of the graviton 192
5.7.1. Canonical quantization 194
5.7.2. The zero mode 201
5.7.3. Path integral quantization 203

**Part 2. Quantum braneworld phenomenology** 205

Chapter 6. Radiative constraints on brane quintessence 207
6.1. Quintessence and dark energy in four dimensions 209
6.2. Quintessence in the braneworld 212
6.3. Radiative corrections to quintessence couplings and masses 213
6.3.1. Constraints on the mass term $m(Q)$ 214
6.4. The gravitational propagator 218
6.4.1. The Einstein–Hilbert action 220
6.4.2. Quantization of the graviton theory 225

6.4.3. The Randall–Sundrum propagator 227
6.5. The brane matter theory 231
6.6. Gravitational coupling of quintessence 233
6.6.1. Loop diagram 233
6.6.2. Triangle diagram 236
6.7. Vacuum polarization and the quintessence mass 240
6.8. Summary 243

Chapter 7. Bulk quantum fields, perturbation theory, and breathing orbifolds 245
7.1. Introduction 245
7.2. Braneworld power spectrum 249
7.3. The consistency relation 253
7.4. Fluctuations in a perturbed four-dimensional de Sitter space 256
7.4.1. Introduction 256
7.4.2. Scalar and tensor power spectra 257
7.5. Fluctuations in a perturbed de Sitter braneworld 262
7.5.1. The tensor zero mode 264
7.5.2. Braneworld consistency relation 268
7.6. Summary 269

Chapter 8. Quantum cosmology of Randall–Sundrum type models 273
8.1. Quantum cosmology in $(3+1)$-dimensions 279
8.1.1. The gravitational Hamiltonian 281
8.1.2. The Wheeler–de Witt equation 283
8.2. Quantum Randall–Sundrum universes 283
8.2.1. The gravitational Hamiltonian 283
8.2.2. Black hole masses in the bulk 286
8.2.3. Quantum representation 290
8.2.4. Solution of the quantum constraints 292
8.3. Quantum Randall–Sundrum universe and conformally coupled scalar matter 296
8.3.1. Hamiltonian for RS gravity and conformal scalar 296
8.3.2. Quantum representation 298
8.3.3. The boundary wavefunction 300

8.4. The bulk gravitational sector 304
8.4.1. Probability density and early universe behaviour 308
8.5. Summary 309

Chapter 9. Winding modes and other exotica 313
9.1. Gravitational winding modes 314
9.2. Winding mode wavefunctions 318
9.2.1. Fitting boundary conditions 322
9.2.2. Eigenvalues and the mass spectrum 324
9.2.3. 1-loop quantum effective action 328
9.3. Asymptotics of the hypergeometric function 335
9.4. Spherically symmetric braneworlds 337

Appendix A. Functional analysis 343
A.1. The Liouville equation 344
A.1.1. The regular Sturm–Liouville problem 345
A.1.1.1. The Poincaré phase plane. Prüfer system 345
A.1.1.2. Liouville equation in Prüfer form 347
A.1.1.3. Completeness of eigenfunctions. Rayleigh's theorem 348
A.1.2. The singular Sturm–Liouville problem and the Sturm–Liouville transform 350
A.2. The path integral 352
A.2.1. Integration on $SL^2(a, b)$ 352
A.2.2. $\zeta$-function regularization 354

Appendix B. Geometry and topology 357
B.1. Differential geometry 358
B.1.1. Exterior algebra 358
B.1.2. Tangent vectors and differential forms 360
B.1.3. Tensors 361
B.1.4. Push-forward and pull-back 362
B.1.5. Integration of $p$-forms 362
B.1.6. Stokes theorem 363
B.2. Metric structures 363
B.2.1. Hodge star 365

B.2.1.1. Inner product of forms. Metric adjoint 366
B.2.2. Manifolds 367
B.3. Homology and cohomology; the de Rham cohomology 367
B.3.1. de Rham cohomology 368
B.3.2. General results for the de Rham cohomology 369
B.3.3. de Rham cohomology ring 370
B.3.4. Simplicial homology 370
B.3.5. Hodge–de Rham theory 371
B.4. A summary of fibre bundles 372
B.4.1. Principal bundles 373
B.4.1.1. Lie groups 373
B.4.2. Bundles of other types 375
B.5. Connexions 375
B.5.1. Parallel transport 376
B.5.2. Field strength and curvature. Cartan structural equations 377
B.6. Einstein–Cartan theory 379
B.6.1. Weyl gravity and conformal compensation of Einstein gravity 380

Appendix C. Squeezed cosmological states and the transition to semiclassical behaviour 383
C.1. Squeezed states and expanding universes 383
C.1.1. Time evolution of $b_{\mathbf{k}}$ 387
C.1.2. Squeezing 389
C.2. Transition to semiclassical behaviour 389

Appendix D. Holography, AdS/CFT and dS/CFT 393
D.1. The renormalization group 393
D.1.1. The Callan–Symanzik equation 394
D.2. Holographic inflation 396

Bibliography 401

Index 423

# Declaration

I declare that this thesis has been composed by myself, and that the work presented is either my own work, or where performed in collaboration I have made a substantial contribution to the work, the extent of such a contribution being indicated in the text.

David Seery

17th December 2004

# Acknowledgements

It is a pleasure to thank those who have helped supervise my work over the last three years. In particular, thanks are due to Andy Taylor (Chapter 7), Bruce Bassett (Chapter 6 and Section 9.4), Alan Heavens, John Peacock. In addition, David Wands, Jim Lidsey and Andrew Liddle offered helpful comments. Portions of this thesis were worked on, and partly facilitated, while enjoying the hospitality of the University of Portsmouth, and while participating in the programme *Prospects in Theoretical Physics – Cosmology, Particles and Strings* at the Institue for Advanced Study, Princeton (30 June – 11 July 2003). I would like to acknowledge financial support from the Institute for Astronomy, the Particle Physics and Astronomy Research Council (PPARC), and a scholarship from the Institute for Advanced Study. This work was completed with the assistance of a PPARC studentship (reference PPA/S/S/2001/03209).

Parts of the original research presented in this thesis have previously appeared in print. The material which constitutes the present Chapter 6 was published in the *Journal of Cosmology and Astroparticle Physics* (Seery and Bassett, 2004). The present Chapter 7 is in press with *Physical Review D* (Seery and Taylor, 2003). A paper based on Chapter 8 has been submitted to *JCAP*.

David Seery

Edinburgh, 2004

# Introduction

This thesis deals with the interaction of quantum mechanical models and cosmologies based on brane universes, an area of active theoretical speculation over the last five years.

For convenience, the material has been split into two parts. Part 1 deals with a selection of background topics which are necessary and relevant to the original research. This research is presented in Part 2. In addition, some auxiliary topics, both more elementary and more advanced, are described in the appendices. The selection of background topics has been influenced by the various techniques, physical theories and mathematical technologies which play a major role in the work presented in Part 2. Although the exposition is *ad hoc*, an attempt has been made to systematically develop portions where the technique (or use of it) may be unfamiliar.

A fairly complete treatment of the necessary mathematical scaffolding is supplied. Although important, this material is familiar or strongly mathematical, and is deferred to the appendices. This includes an elementary survey of functional analysis in Appendix A, sufficient to support a discussion of the path integral. The path integral formalism is used extensively throughout this thesis, and, where available, constitutes our preferred representation of quantum mechanics. The discussion is limited to the relevant portions of the theory: functions in Banach spaces, and the Sturm–Liouville basis (technology which appears many times in Part 2); direct evaluation of Gaussian functional integrals, ubiquitous in field theory calculations; and $\zeta$-function regularization of the operator determinants to which such Gaussian integrals give rise, which has a direct application in Chapter 9. In Appendix B we describe the necessary framework of differential geometry which supports general relativity, and low-energy discussions of string theory. All calculations in metric gravity are based on differential geometry, together with a good proportion of the technology which buttresses quantum field theory on curved space time, string theory, and some more advanced representations of quantum mechanics (see below). All of this is used extensively throughout both parts of the thesis. We include some more advanced topological

technology which supports the discussion of string compactification. General results from compactification theory, when appropriately interpreted in the brane context, contribute important stability results for zero-modes of the Kaluza–Klein fields, and provide a natural home for the spectral KK technology used (in one form or another) throughout Part 2, but most especially in Chapter 7 and Chapter 8. Einstein gravity and Yang–Mills theory are set in context as examples of connexions on fibre bundles.

Part 1 opens with a brief recapitulation of general relativity in its elementary formulation, which can be contrasted with the Cartan theory of Appendix B. This is the basic technology, familiar from every undergraduate physics course, which we rely on everywhere. We make little of use of advanced toplogical or global methods in gravity, so the discussion can be limited to the mere calculation of curvature quantities and the Einstein field equations. We supply details of the Lagrangian formulation, and discuss the decomposition of curvature in situations where one or more distinguished hypersurfaces are present in the spacetime. Apart from these everyday and effective tools, there is very little that we need to borrow from the vast technical literature of metric gravity.

Chapter 2 takes up the issue of quantum mechanics, and quantum field theory in particular, relying on the general mathematical framework established in Appendix A and the tools of geometry and gauge theory outlined in Appendix B. Quantum mechanics is discussed in a fairly general setting, and we attempt to avoid a sterile presentation of the undergraduate-textbook theory, preferring instead to work with representations of states and observables on phase space. Quantum and classical mechanics are presented as distinct representations of the same theory. We briefly allude to even more generalized frameworks in which quantum mechanics can be understood as an algebraic deformation of the classical theory. Finally we discuss quantum gauge field theory and theories with constraints, spending some time on a derivation of the Fadeev–Popov determinant. These considerations are important for the later discussion of interacting quantum fields on the brane, and quantum Randall–Sundrum universes which are both described by gauge-invariant field theories. As a side issue, not directly related to the work which will be described in Part 2, the BRST quantization of rather general field theories is described. This completes the Fadeev–Popov story, and introduces tools and ideas which are important for the quantization of the Polyakov string in the next chapter.

Chapter 3 deals with a few selected issues in string theory. Despite the heavy interdependence of string theory, M-theory and brane compactifications – which provide the broadest theoretical underpinnings for brane universes, and endow them with their rich physical phenomenlogy – the subject of string theory *per se* is mostly tangential to the models discussed in this thesis. The familiar and universal tools of general relativity and quantum field theory are mostly sufficient to discuss the truncated, low-energy versions of string theory which are calculationally accessible in our models, although the restriction to these methods does somewhat limit what can be said at high energy or strong coupling. Instead, the limited discussion of string theory presented in this chapter is only intended to provide a context for the particular models we shall employ, and to explain how they fit into the much larger framework of string compactifications. We introduce the Polyakov string and discuss its gauge fixing, before solving heuristically for the massless excitations in the string spectrum. At this point, we depart from the conventional presentation of the quantum relativistic string and move rapidly to the subject of string compactifications, usually reserved as a more advanced topic. There is not much necessity to describe the sophisticated topological machinery which is needed to handle fields of arbitrary spin: we are only dealing with bosonic fields which for practical purposes have Poincaré spin zero, so the de Rham cohomology, as outlined in Appendix B, is quite sufficient and there is no justification for a painful passage to the Atiyah–Singer theorem and the Dirac index. Instead, we derive the connexion between the number of zero-modes of a given $p$-form and the Betti number $b_{n-p}$ (where $n$ is the dimension of spacetime) of the compactification manifold just to the point where the result is plausible, and indicate how the stability of zero modes is a *topological* concept. The remaining issues which must be dealt with, from a stringy perspective, are M-theory and the string dualities. We introduce T-duality via the exchange of winding and momentum modes on a torus, and explain how this leads to the appearance of objects with Dirichlet boundary conditions in the theory, conventionally called D-branes. Finally we discuss a little of the phenomenology of M-theory compactifications, concentrating on how cosmological scenarios may emerge from a sector of moduli space near the weakly coupled $E_8 \times E_8$ heterotic string. This will be pursued in much greater detail in Chapter 5.

In Chapter 4 we justify our focus on low energy phenomenology by briefly describing the current state of observational cosmology, paying particular attention to the cosmic

microwave background experiments which are most relevant to speculations about high energy physics. The present emergence of rather more precise experiments and data sets than has customarily been available to cosmologists makes attempts to bridge the gap between the conservative, but empirically solid four-dimensional classical cosmology, and radical (but theoretically well-founded) speculations about physics at or near the string, GUT or Planck scales most timely. For this reason we identify the major epochs in the lifecycle of the universe, until the present moment, and enter into some detail concerning the gross thermal history of the universe, the growth of perturbations which will ultimately collapse into the matter haloes hosting our own and other galaxies, and the gross details of the production of cosmic microwave background anisotropies. In any case, all of this is necessary to understand the motivations for the esoteric and seemingly baroque mechanisms invoked freely in Part 2, which all have their origin in puzzles and conundrums which arise out of our understanding of quantitative, predictive cosmology and its secure foundation in accelerator physics. We briefly discuss the most outstanding of these issues, the horizon and flatness problems, and introduce a central concept: early universe inflation. The idea of inflation motivates a good deal of the original research presented here, but comes with its own difficulties and perplexities which it is part of the purpose of the braneworld scenario to unravel. All of this material is purely background, but is assumed implicitly in Part 2. We discuss the quantum field theoretic calculations which seem to show that inflation can provide a causal mechanism to seed structure formation in the early universe (technology which is absolutely vital later on) and conclude with a brief summary of the observational position.

An alternative view of inflation is sketched in Appendix D, which also provides a brief summary of the holographic principle which is occasionally invoked in the text. This material does not constitute an important part of the work and can be read independently or skipped if desired. In addition, we summarise the theory of the quantum-to-classical transition for the inflationary perturbation spectrum in Appendix C which is based on the idea of quantum squeezing from the expansion of the universe. This is, in fact, a central pillar of the inflationary model, and although it plays no part in our explicit calculations it seemed entirely unreasonable to omit it.

Having condensed sufficient string theory and elementary cosmology to inform the discussion, we provide a mini-review of some selected aspects of brane cosmology, from the

perspective which will be important in Part 2. This constitutes Chapter 5. We introduce the necessary geometrical constructions and solve for the Binetruy–Deffayet–Ellwanger–Langlois metric before presenting the two canonical analytic examples which we shall employ extensively, Randall–Sundrum (Minkowski) branes and Kaloper–Linde (de Sitter) branes. Despite their importance, these are by no means the only braneworld models. Two other important classes are the direct heterotic M-theory compactifications, when appropriately truncated to the massless modes which will be important in four dimensions, and the Ekpyrotic and cyclic scenarios occasionally advocated as providing explanations for some of the most puzzling and troublesome features of the Standard Model. Both these families of models allow for the possibility of colliding branes. The discussion of brane phenomenology to appear in Part 2 concerns only the BDEL metrics, but a diversity of models offers a larger context and supplies a method of critical comparison. Some brief comments about global methods can be sketched, beginning with a derivation of the effective Einstein equations on the brane, which makes transparent how traditional four-dimensional Einstein gravity is modified in the brane scenario. After discussing stability issues (something of a problem in the general brane compactification) we explain a construction due to Verlinde which provides an AdS/CFT interpretation of the general braneworld compactification and supplies valuable insight into the general behaviour of brane geometries, before moving on to the Kaluza–Klein theory of the graviton, which is secretly nothing other than the Kaluza–Klein theory of a spin zero field. This concludes the presentation of basic, necessary background material.

Part 2 is concerned with phenomenological tests, and opens with a direct calculation in Chapter 6. The aim here is to use stringent limits on the mass of a certain class of scalar fields, including the inflaton but also any putative scalar fields which might drive the late-time acceleration of the universe which we are currently observing, to limit the parameters of brane cosmological models. There are certain severe bounds on the possible gravitationally-mediated couplings of such scalar fields to conventional matter, such as Standard Model fermions like quarks or leptons. The major theoretical technology assembled in this chapter is a proper derivation of the gauge-fixed graviton action, which we accomplish by the Fadeev–Popov procedure, and an examination of the Feynman rules which couple an on-brane quantum field theory to the bulk gravitational sector. These rules can be read off from the gauge-fixed action, when supplemented with appropriate

fermion and scalar fields. For simplicity we work with Minkowski branes and focus on the phenomenological consequences of a low controlling Planck scale. The most sensitive probe of a low fundamental scale in the low-energy field theory will be loop processes in which the effective momentum cut-off is drastically reduced from the four-dimensional scale of $10^{19}$ GeV or so to scales around the SUSY mass, possibly at 1 TeV. On the basis of these one-loop calculations we are able to draw some careful conclusions about the viability of inflationary epochs, such as early universe inflation and late-time acceleration, on the brane.

Chapter 7 is concerned with a different experimental signature, namely the appearance of the cosmic microwave sky. The calculation of the spectrum of gravitational waves in the braneworld can now be considered classical, and was presented in Chapter 5. After discussing the observational situation, we point out an observation made originally by Huey & Lidsey, that it is observationally difficult (and potentially impossible) to distinguish brane inflation from conventional inflation. This is potentially damaging for the kind of phenomenology we are interested in: if the four-dimensional world is observationally indistinguishable, there is no conceptual need at all for the extra epicycles introduced by brane mechanisms, and the formalism would be relegated to a mere curiosity (provided it does not contribute more meaningful understanding elsewhere, naturally). We describe a construction which suggests that when higher order quantum effects are taken into account there may be distinguishable characteristics of brane inflation. This endeavour necessitates the introduction of some new machinery to handle compactifications which include a breathing mode of the transverse dimension. This is done perturbatively over the top of the background Kaloper–Linde brane and relies on a path-integral calculation of the zero-mode spectrum. Unfortunately, the various elegant tools which were available in simpler cases where the spacetime can be given a bundle structure as a fibration of spacetime over the compact dimension cannot be applied here. This toplogical obstruction is the cause of the difficulty in solving for the behaviour of gravity on generally disturbed orbifolds. Instead we develop an ad hoc perturbation theory which works for the zero mode, but unfortunately we cannot bootstrap our result to say anything about higher Kaluza–Klein modes.

Having studied the behaviour of quantum fields propagating over the brane compactification, both in the case where such fields are attached to the brane, corresponding to open

string modes in the string-theory picture, and quantum fields in the bulk, which are closed strong modes, we turn to a quantization of the brane compactification itself. The full M-theory compactification is presumably a quantum mechanical affair, and the low-energy classical backgrounds which are are describing do not capture the quantum behaviour. Proper control over the quantum aspects can only properly be arrived at by considering the M-theory model, but as a compromise there is a good deal of sense in quantizing the general relativity model. This could not be expected to provide details concerning the model's behaviour in regions near the string scale, but its qualitative aspects should be trustworthy in energy régimes where the typical excitations have lengthscales somewhat longer than the Planck length. By comparing the results with the well-established theory of quantum cosmology in four dimensions we can begin to understand whether and if the brane picture makes different predictions about the earliest epochs and the possible beginning of the universe via quantum processes. This subject is ripe for investigation and has been addressed previously in the literature, but we provide the first five dimensional picture that takes into account the full five dimensional action.

In a final chapter, we collect some miscellaneous calculations which have not yet led to firm predictions. The first of these is an attempt to introduce winding modes of the extra transverse dimension into the spectrum. This is fairly straightforward in the case of Minkowski branes but suffers from some technical complexity in the case of Kaloper–Linde branes. We use a recently introduced $\zeta$-function technique to sum the one-loop effective action and look for interesting behaviour as the branes approach each other. In models where brane collisions play a major role, one can expect that strings stretched between the branes which are becoming light near the collision are heavy when are far away, and vice versa. This observation is the basis of an important formal property of string theory named T-duality. The truncation of winding modes from the low-energy spectra may mean that the two opposite régimes of near- and far-branes have to be treated separately. The second calculation concerns the appearance of Birkhoff's theorem in the braneworld. In this case we are only able to show that the reasons for its emergence in four dimensions do not translate easily to the brane world, but the calculation is intended to be a starting point for the investigation of black holes on the brane. Astrophysical black holes are now a well-established fact, and differing properties in the braneworld could provide an opening to study strong field gravity on the brane: for example, there is no Carter–Robinson theorem

in five dimensions. The zoology of braneworld black holes can therefore be expected to be somewhat more diverse than in four dimensions.

# Part 1

# Background: Cosmology and Quantum Mechanics

CHAPTER 1

# Low energy physics: notation and conventions

The physical theories involved in the description of the kinds of cosmological model that this thesis is concerned with can be circumscribed almost entirely by the two pillars of the modern theoretical world, Yang–Mills theory and Einstein gravity. In order to provide an appropriate framework for the more exotic objects and excitations which are seen in these high energy models we will have to add some extra superstructure later, namely string theory and its conjectural covering theory, M-theory. Nonetheless, all the calculations to be described in Part 2 rely only on the conventional, familiar tools of gauge theory and gravity.

In this chapter, we offer a brief outline of the calculational aspects of the necessary technologies, purely at the classical level. The mathematical elements are founded to a large degree on the tools of differential geometry which are supplied by the mathematicians; these elements are reviewed in Appendix B. Additionally, we are not concerned here with quantum effects, the process of quantization, or quantum field theory in general, which will properly be the subject of Chapter 2. Instead we are interested in exhibiting those techniques and calculational methods which will be most relevant for our applications. In general relativity, this means the formalism of curvature in the Einstein holonomic gauge and the $(n+1)$ description of curvature (but see Section 1.2 below). Brane universes do not pose a sufficient calculational problem that the extra machinery involved in coordinate-free calculations is worthwhile, and $(n+1)$ decomposition of the curvature quantities provides a powerful and elegant method of handling codimension one branes of the sort which will dominate the later discussion.

In addition, we recount here sufficient of the theory of Clifford algebras to handle the occasional use we shall make of spinors (Section 1.2). Unfortunately this subject, and indeed general relativity as a whole, suffers from a proliferation of mutually incompatible and contradictory sign and notational conventions, which engenders a particular responsibility to spell out with some explicitness which choices are being used. The present chapter is

also intended to discharge this responsibility, and summarise the notation which will be used uniformly throughout the rest of this thesis.

## 1.1. Einstein gravity

There are many routes to discovering Einstein gravity. One can begin with the formalism of connexions on a $GL_4$ gauge bundle, as we do in Appendix B, and describe the curvature tensor by a field strength on the bundle which is subject to the Cartan structural equations. Alternatively one can try and introduce gravity as the gauge theory of the Lorentz group (Chamseddine and West, 1977), which is an approach that leads to a natural and useful generalization to the construction of supergravity (Freund, 1988). The more familiar physicist's approach was introduced by Weinberg, who appealed to general covariance as a bootstrapping technique to obtain general relativity from special relativity (Weinberg, 1972). This approach has admirable directness and an appealing transparency, but is subject to coupling ambiguities (see, eg., Carroll (1997)). On the other hand, seeking mathematical clarity, one can appeal to the simplicity of the action principle, as was first done by Hilbert, in order to derive the lowest-order Lorentz-invariant field equations. These equations can be verified, or developed constructively *ab initio* by requiring that they obey sensible physical conditions such as covariant conservation (Lovelock, 1971). In four dimensions, the answer one finds is unique:

$$G_{ab} = \kappa^2 T_{ab}, \tag{1.1.1}$$

where $G_{ab}$ is a second-rank symmetric tensor built out of a particular combination of curvature invariants,

$$G_{ab} = R_{ab} - \frac{1}{2} R g_{ab}, \tag{1.1.2}$$

and this equation is supposed the govern the curvature of spacetime, which is a differential manifold carrying a metric $g_{ab}$ and a natural notion of parallel transport which is compatible with the metric. These concepts are summarised in Appendix B. One can show (Weinberg, 1972) that $G_{ab}$ obeys the conservation law $\nabla^a G_{ab} = 0$, called the Bianchi identity, given that $\nabla_a$ is the covariant derivative associated with the natural metric notion of parallel transport on spacetime (Appendix B). The Ricci tensor $R_{ab}$ is the contraction of the Riemann curvature tensor, which describes how covariant derivatives commute,

$$[\nabla_a, \nabla_b] V_c = R_{abcd} V^d \quad \text{(for all } V_c\text{)}. \tag{1.1.3}$$

There are a variety of more or less messy constructions which allow one to relate $R_{abcd}$ to the curvature of the manifold, but this definition is by far the cleanest and most natural. Let $\Gamma$ be the metric connexion. Then, $R^a{}_{bcd}$ is described in components by the rule

$$R^a{}_{bcd} = \partial_c \Gamma^a_{bd} - \partial_d \Gamma^a_{bc} + \Gamma^f_{bd}\Gamma^a_{fc} - \Gamma^f_{bc}\Gamma^a_{fd}. \tag{1.1.4}$$

We define the Ricci tensor to be the contraction $R_{bd} = R^a{}_{bad}$ and the Ricci scalar to be $R = \operatorname{Tr} R = R^a{}_a$. The connexion $\Gamma$ can be written in terms of the metric in the form

$$\Gamma^a_{bc} = \frac{1}{2} g^{ad}(\partial_b g_{dc} + \partial_c g_{bd} - \partial_d g_{bc}). \tag{1.1.5}$$

**1.1.1. Coupling to matter.** One can couple this theory to arbitrary matter either by returning to the minimal action principle and writing all Lagrangians as sensibly formulated invariant integrals over spacetime, after which curvature components and factors of the metric will appear in all matter field equations and the energy–momentum tensor $T_{ab}$ in a correct way to couple matter to gravity, or by promoting the known matter field equations to covariant form, in which all derivatives are replaced by covariant derivatives and equations are restricted to those which can be cast in tensorial form. Such equations are invariant under diffeomorphisms which mix the coordinates of the manifold, as expected for sensible physical laws. In particular, the worldlines of freely falling observers with tangent vector $u^a$ are subject to the free-fall equation $u^a \nabla_a u^b = 0$, which expresses conservation of their four-velocity. This is to be compared with the unaccelerated expansion law $\dot{v} = 0$ which is familiar from Newtonian dynamics.

Particles arrange their motion in such a way to minimise the proper length of their worldlines, so the action principle is

$$S = \int_{\text{worldline}} \mathrm{d}s, \tag{1.1.6}$$

where $\mathrm{d}s$ is the infinitesimal proper displacement caused by any given coordinate displacement, and can be expressed once coordinates are chosen by the law

$$\mathrm{d}s^2 = g_{ab}\mathrm{d}x^a \mathrm{d}x^b. \tag{1.1.7}$$

Since it is only the length of the worldline which is important, any monotonic function of $\mathrm{d}s$ will do equally well, and it is frequently convenient to work with the action principle in the form $\int \mathrm{d}s^2$ instead.

Couplings to fields, rather than particles, can be accommodated equally easily by writing the action as an invariant integral and computing the energy–momentum tensor $T_{ab}$, which can be written in terms of the matter Lagrangian density $\mathcal{L}$ as

$$T_{ab} = -\frac{2}{\sqrt{-g}}\frac{\delta \mathcal{L}}{\delta g^{ab}}. \tag{1.1.8}$$

In virtue of the Noether theorem this tensor is automatically covariantly conserved, $\nabla^a T_{ab} = 0$, so it is a natural candidate to appear on the right-hand side of the Einstein equation (1.1.1) after the Bianchi identities have been enforced.

## 1.2. Spinors

We use spinors when dealing with fermionic matter in field theory, in discussing the superstring, and when outlining the features of low-energy $\mathcal{N} = 1$ supersymmetry. Terminology in the spinor world can rapidly become opaque, so we provide a brief summary. (The reader should consult any of Figueroa-O'Farrill (2001); Freund (1988); Galperin, Ivanov, Ogievetsky, and Sokatchev (2001); Göckeler and Schücker (1987); Stewart (1991); Weinberg (1994) for more details.)

Let $\mu$, $\nu$, ... describe indices transforming under the Lorentz group. In the vierbein formalism, the metric $g_{ab}$ (with $a$, $b$, ... describing coordinate indices which transform under the group of coordinate diffeomorphisms, or $GL(n)$ in $n$ dimensions) is described by a vierbein field $e^{\mu}_{a}$,

$$g_{ab} = \eta_{\mu\nu} e^{\mu}_{(a} e^{\nu}_{b)}. \tag{1.2.1}$$

One should consider the $e^{\mu}_{a}$ to be valued in the gauge group $SO(1, n-1)$, or, since we are dealing with spinors, more properly its universal covering group $Spin(1,3)$ (Göckeler and Schücker, 1987). The $e^{\mu}_{a}$ perform the change of basis necessary to carry a representation of $Spin(1,3)$ from the Einstein gauge, in which the basis is holonomic, to the spinor case, where the basis is orthonormal up to signature. The Dirac algebra is

$$\{\gamma_\mu, \gamma_\nu\} = 2\eta_{\mu\nu}. \tag{1.2.2}$$

The matrices $\gamma_\mu$ are called Dirac matrices. Their indices can be projected onto spacetime using the vierbein via the usual rule, $\gamma_a = e^{\mu}_{a}\gamma_\mu$. As is customary in the physics literature, we omit the identity element of the Clifford algebra when writing the right hand side of the Dirac anticommutation rule.

Let $\Lambda^{a'}{}_{a}$ be an infinitesimal Lorentz transformation, such that when acting on the coordinate fields $x^a$,

$$x^a \mapsto x^{a'} = \Lambda^{a'}{}_{a} x^a = (\delta^{a'}_{a} + \omega^{a'}{}_{a}) x^a, \tag{1.2.3}$$

where $|\omega^{a'}{}_{a}| \ll 1$ and $\omega_{ab} = -\omega_{ba}$. A field in a general representation of the Lorentz group transforms according to the law

$$T^{\mu\nu\cdots\sigma} \mapsto T^{\mu'\nu'\cdots\sigma'} = [U(\Lambda)]^{\mu'\nu'\cdots\sigma'}{}_{\mu\nu\cdots\sigma} T^{\mu\nu\cdots\sigma}, \tag{1.2.4}$$

where the operator $U(\Lambda)$ implementing Lorentz transformations on $T^{\mu\nu\cdots\sigma}$ satisfies

$$U(\Lambda) = 1 + \frac{1}{2}\omega^{ab}\Sigma_{ab}, \tag{1.2.5}$$

and takes values in the representation of $T$ (indices suppressed). The generators $\Sigma_{ab}$ obey the commutation law

$$\begin{aligned} [A_i, A_j] &= \mathrm{i}\varepsilon_{ijk} A_k \\ [B_i, B_j] &= \mathrm{i}\varepsilon_{ijk} B_k \\ [A_i, B_j] &= 0, \end{aligned} \tag{1.2.6}$$

where

$$\begin{aligned} A_i &= \frac{1}{2}\left(-\mathrm{i}\varepsilon_i{}^{jk}\Sigma_{jk} + \Sigma_{i0}\right) \\ B_i &= \frac{1}{2}\left(-\mathrm{i}\varepsilon_i{}^{jk}\Sigma_{jk} - \Sigma_{i0}\right). \end{aligned} \tag{1.2.7}$$

These commutation relations just constitute two copies of the angular momentum algebra, so any representation of the Lorentz group can be decomposed into two representations of the angular momentum algebra of spin $A$, $B$. Such representations are written $(A, B)$.

Consider a representation of $(\frac{1}{2}, 0)$ in an even $d$-dimensional spacetime. Such a representation can be considered as a $2^{d/2-1}$-dimensional symplectic vector space $S$ over $\mathbf{C}$, dual to a space $S^*$, where the isomorphism between $S$ and $S^*$ is induced by the symplectic structure,

$$S^* \ni \kappa_A = \varepsilon_{BA}\kappa^B = \kappa^B \varepsilon_{BA} \in S. \tag{1.2.8}$$

The index raising operator $\varepsilon^{AB}$ is numerically equal to $\varepsilon_{AB}$,

$$\varepsilon^{AB} = -(\varepsilon^{-1})^{AB}. \tag{1.2.9}$$

With the sign chosen in this way, an $S$ vector $\kappa^A$ satisfies $\kappa^A = \varepsilon^{AB}\kappa_B$. Because $S$ is a vector space over $\mathbf{C}$, one might naïvely assume that the complex conjugate of a spinor

in $S$ could be defined as an element of $S$. However, if $\alpha, \beta \in S$ and $c \in \mathbf{C}$, then the complex conjugate of $\alpha + c\beta$ would be $\bar{\alpha} + \bar{c}\bar{\beta}$, not $\bar{\alpha} + c\bar{\beta}$, so complex conjugation is an anti-isomorphism (and not an isomorphism) from $S$ to another vector space $\bar{S}$, which corresponds to the $(0, \frac{1}{2})$ representation. Spinors in $\bar{S}$ acquire a bar, and a dot on the spinor index to distinguish them from elements of $S$. Elements of $S$ or $\bar{S}$ are called Weyl spinors, or sometimes chiral spinors.

A Dirac spinor in $d$ dimensions can be considered as an element $\psi^\alpha$ of a $2^{d/2}$-dimensional vector space isomorphic to $S \oplus \bar{S}^*$. If $\kappa^A$ and $\mu^A$ are in $S$, we can write

$$\psi^\alpha = \begin{pmatrix} \kappa^A \\ \\ \bar{\mu}_{\dot{A}} \end{pmatrix}. \tag{1.2.10}$$

In this notation, it is easy to see that the Dirac representation is the direct sum $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$. The dual space $D^*$ is obviously isomorphic to $S^* \oplus \bar{S}$, and there are two natural maps from $D \to D^*$. The first is the Dirac adjoint,

$$\psi^\alpha \mapsto \bar{\psi}_\alpha = \begin{pmatrix} \mu_A & \bar{\kappa}^{\dot{A}} \end{pmatrix}, \tag{1.2.11}$$

while the other possibility is Majorana conjugation,

$$\psi^\alpha \mapsto (\psi_M)_\alpha = \begin{pmatrix} \kappa_A & \bar{\mu}^{\dot{A}} \end{pmatrix}. \tag{1.2.12}$$

The Dirac inner product is

$$\bar{\psi}\psi = 2\,\mathrm{Re}(\mu_A \kappa^A). \tag{1.2.13}$$

A Majorana spinor is one for which $\psi_M = \bar{\psi}$, which requires $\kappa^A = \mu^A$.

Define $\gamma_{ab}$ by the commutator

$$\gamma_{ab} = \frac{1}{2}[\gamma_a, \gamma_b]. \tag{1.2.14}$$

When acting on a spinor field, the covariant derivative is defined in terms of the spin connection $\omega_a^{\mu\nu}$,

$$\nabla_a \psi = \partial_a \psi + \frac{1}{8}\omega_a^{\mu\nu}\gamma_{\mu\nu}\psi, \tag{1.2.15}$$

where $\omega_a^{\mu\nu}$ is given by

$$\omega_a^{\mu\nu} = -e_b^\nu \partial_a e^{b\mu} - \Gamma_{ac}^b e^{c\mu} e_b^\nu. \tag{1.2.16}$$

One can build a curvature out of the spin connection in the usual way,

$$R_{\mu\nu ab} = \partial_\mu \omega_{\nu ab} - \partial_\nu \omega_{\mu ab} + \omega_{\mu a}^c \omega_{\nu cb} - \omega_{\mu b}^c \omega_{\nu ca}, \tag{1.2.17}$$

where $\partial_\mu = e^a_\mu \partial_a$. By projecting tangent space indices back onto spacetime, one recovers the usual Riemann curvature. The spinor Ricci identity is

$$[\nabla_a, \nabla_b]\psi = \frac{1}{8} R_{abcd}\gamma^{cd}\psi. \tag{1.2.18}$$

This expression for the curvature makes it clear that the Riemann 'tensor' actually contains two distinct types of indices (see Section B.5.2)

## 1.3. The $(n+2)$ description of curvature

In circumstances where a distinguished hypersurface exists, one can use what is known as the $(n+1)$-decomposition of the curvature components. This calculation is described in detail in Hawking and Ellis (1973); Visser (1996). In the case which will be of interest later, it is possible to distinguish two possible codimension one surfaces, which are hypersurfaces on which $t$ = constant, with timelike unit normal $t_a$, and hypersurfaces for which $y$ = constant with spacelike unit normal $y_a$. We give the decomposition of curvature taking into account both these hypersurfaces, which can then be specialised to use just one or both, as required. Although this technology may have been applied before, it is not widely reported in the literature and was developed *ab initio* for the purposes of this chapter. One defines a projection tensor $h_{ab}$ which is orthogonal to both these hypersurfaces,

$$h_{ab} = g_{ab} + t_a t_b - y_a y_b. \tag{1.3.1}$$

This tensor is a projection, in the sense that

$$h_a{}^b h_b{}^c = h_a{}^c \tag{1.3.2}$$

$$h_a{}^b t_b = h_a{}^b y_b = 0 \tag{1.3.3}$$

and $h_a{}^b$ projects into the three-dimensional subspace orthogonal to $t_a$ and $y_a$. One defines a covariant derivative in this space by projection,

$$\overset{n}{\nabla}_a Y_{b\ldots} = h_a{}^{a'} h_b{}^{b'} \overset{3}{\nabla}_{a'} Y_{b'\ldots}. \tag{1.3.4}$$

One can then build a $GL_n$ curvature via the Ricci rule[1]

$$\overset{n}{R}_{abcd} Y^d = [\overset{3}{\nabla}_a, \overset{3}{\nabla}_b] Y_c. \tag{1.3.5}$$

[1]This is the same procedure that one employs in standard Riemannian geometry to build a curvature, so there is nothing special about the technique. The only real novelty here is that the $GL_3$ curvature itself can be expressed in terms of a more fundamental tensor field, the $(n+2)$ dimensional curvature tensor.

The aim is to express the full five-dimensional curvature $R$ in terms of the $GL_n$ curvature $\overset{n}{R}$ and the intrinsic and extrinsic curvatures of the hypersurfaces $t$ = constant and $y$ = constant. What follows is a summary of the fairly lengthy calculations involved in making this programme concrete.

Define a spatial extrinsic curvature $\psi_{ab}$ and a timelike extrinsic curvature $\Omega_{ab}$ via

$$\psi_{ab} = h_a{}^{a'} h_b{}^{b'} \nabla_a t_b \quad \text{and} \quad \Omega_{ab} = h_a{}^{a'} h_b{}^{b'} \nabla_a y_b. \tag{1.3.6}$$

One finds that $\overset{n}{R}$ can be expressed in terms of $R$ and invariants constucted from $\psi_{ab}$ and $\Omega_{ab}$,

$$\overset{n}{R} = R + 2t^a t^b R_{ab} - 2y^a y^b R_{ab} - 2t^a t^c y^b y^d R_{abcd} - \psi^2 + \Omega^2 + \psi^{ab}\psi_{ab} - \Omega^{ab}\Omega_{ab}, \tag{1.3.7}$$

where $\psi = \operatorname{Tr} \psi_{ab}$ and $\Omega = \operatorname{Tr} \Omega_{ab}$, and the trace can be constructed either from $g_{ab}$ or $h_{ab}$. To find $R_{ab}t^a t^b$ and $R_{ab}y^a y^b$, one can use a trick based on the Ricci identity,

$$[\nabla_a, \nabla_b] t_c = R_{abcd} t^d \quad \text{so} \quad t^b [\nabla_a, \nabla_b] t^a = R_{bd} t^b t^d. \tag{1.3.8}$$

This is Codacci's equation (Hawking and Ellis, 1973; Visser, 1996). Proceeding in this way, one finds

$$R_{ab}t^a t^b = \nabla_a \varphi^a - \psi^{ab}\psi_{ab} + \psi^2 + (h^{ab}y^c y^d + h^{cd}y^a y^b - h^{ac}y^b y^d - h^{bd}y^a y^c)\nabla_a t_b \nabla_d t_c \tag{1.3.9}$$

where $\varphi^a = t^b \nabla_b t^a - t^a \nabla_b t^b$; and

$$R_{ab}y^a y^b = \nabla_a \omega^a - \Omega^{ab}\Omega_{ab} + \Omega^2 - (h^{ac}t^b t^d + h^{bd}t^a t^c - h^{ab}t^c t^d - h^{cd}t^a t^b)\nabla a y_b \nabla_d y_c \tag{1.3.10}$$

where $\omega^a = y^b \nabla_b y_a - y^a \nabla_b y^b$. Collecting terms, and taking appropriate care about cancellations, one can show that

$$\begin{aligned}\overset{n}{R} = R &+ 2\nabla_a(\varphi^a - \omega^a) + \psi^2 - \psi^{ab}\psi_{ab} + \Omega^{ab}\Omega_{ab} - \Omega^2 \\ &- 2t^a t^c y^b y^d R_{abcd} - 4y^b y^d \nabla_a t_b \nabla_d t^a + 4y^a y^b \nabla_a t_b \nabla_d t^d - 4t^b t^d \nabla_a y_b \nabla_d y^a + 4t^a t^b \nabla_a y_b \nabla_d y^d\end{aligned} \tag{1.3.11}$$

In the $(n+1)$-formalism, this would correspond to Gauss' equation (Hawking and Ellis, 1973; Visser, 1996), for which reason we shall refer to it as the generalized Gauss' equation. This describes $R$ up to a total derivative, which will vanish when $R$ is integrated to form the Einstein action, assuming appropriate boundary conditions on the fields, or at most will give a boundary term which determines how the fields couple to the brane. We shall ignore any such terms for the time being.

CHAPTER 2

# Quantum field theory

This chapter is an introduction to the quantum field theory we shall find it necessary to employ in subsequent chapters. It is intended to be a working account, in the sense that it constitutes a toolbox of techniques on which we shall draw heavily in later parts of the discussion.

As a community, our approach to both the teaching and professional use of quantum theory over the last 80 years or so, and quantum field theory in particular, has probably changed more than any other part of the standard canon of physics. In the early, explorative days (the "old fashioned" quantum theory) both the rigorous underpinnings of the subject and the sophisticated calculational techniques such as Feynman diagrams that we now take for granted were lacking. Therefore, it made sense to begin as far down the classical road as possible, emphasis being given particularly to the wave equations of classical physics and their reinterpretation within the quantum framework. With this background it was common to think of two phases of quantization: the passage from point-particle classical mechanics to the quantum mechanics of single particles, or so-called first quantization, and then the next step of multi-particle quantum mechanics, so-called second quantization. This second step was necessitated by attempted unifications with relativity, in which the classical guarantee that one-particle states remain one-particle states is revoked, and was at first interpreted as a quantization of the particle wavefunctions which had arisen from first quantization. This point of view is reflected in the naming.

These days, we do not approach quantum field theory in the same spirit. The old intepretation of second quantization was abandoned early and the importance of wave equations has steadily dwindled in the face of new ideas for constructing quantum fields ab initio (Weinberg, 1994). Many modern treatments in textbooks and monographs take account of this shifted point of view. The present account is drawn from several of these, and makes no claim of originality. It is highly compressed and the reader is referred to the original source material for more leisurely accounts. Our notation usually coincides with

Weinberg (1994). However, in the present case we aim at the most economical treatment possible, and many of the details set out here are rather closer to Fadeev (1999) and Witten (1999b). We make some use of the phase space formulation in terms of cotangent bundles, for which relevant details can be found in Wald (1994) and Deligne et al. (1999). We do not make much use of the canonical formulation. This is not because of any real inferiority of the canonical theory, but rather because the path integral formulation handles the kind of physics we shall be dealing with (gauge field theories, topological field theories and Kaluza–Klein theories) rather more easily than the canonical approach. A brief explanation of the path integral was given in Appendix A. More details can be found in the monograph of Rivers (1988), or the early standard reference by Feynman and Hibbs (1965), and the subject is covered is many modern general treatments (Peskin and Schroeder, 1995; Weinberg, 1994). Among the most complete and insightful references to appear in recent years is the two-volume Institute of Advanced Study publication by Deligne et al. (1999). Other parts of the treatment are based on ideas and notation outlined in Albeverio, Jost, Paycha, and Scarlatti (1997); Carlip (1998); D'Eath (1996); Galperin et al. (2001); Green, Schwarz, and Witten (1987); Polchinski (1998).

## 2.1. Classical mechanics and classical field theory

Before going over to the quantum world, it is useful to take a step backwards and define what we mean by a classical theory, such as the Newtonian point particle or Maxwell's electrodynamics. It is almost always sufficient to describe a classical theory in terms a number $N$ of variables $q^1, \ldots, q^N$ which typically correspond to positions or angles in configuration space, and in fact all the classical theories we shall encounter here can be described in this way. Mechanics as usually studied in elementary treatments involves the solution of $N$ second-order differential equations for these variables; such a system is said to be an $N$-dimensional theory. Alternatively, one can introduce canonical momenta $p_1, \ldots, p_N$ which are related to the first derivatives of the $q^i$. The equations of mechanics can be rewritten as first order equations in the $2N$ variables $q^i$, $p_j$; such a system is said to be written in Hamiltonian form. The space spanned by the $q^i$, $p_j$ is called phase space and written $M$. Often $M$ is isomorphic to $\mathbf{R}^{2N}$ but this is not necessary and there are examples of toplogically non-trivial phase spaces. For example, in the description of spin phase space is the sphere $\mathbf{S}^2$.

Evolution over $M$ is established by prescribing a Hamiltonian, which is a function $H(p,q)$ on phase space. The dynamics of $q^i$ and $p_j$ are governed by Hamilton's equations,

$$\frac{\mathrm{d}q^i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q^i}. \tag{2.1.1}$$

Write $y_a = (q^i, p_j)$ and introduce the $2N \times 2N$ two-form $\omega_{ab}$ which is chosen appropriately so that Hamilton's equations become

$$\frac{\mathrm{d}y_a}{\mathrm{d}t} = \omega_{ab}\frac{\partial H}{\partial y_b}. \tag{2.1.2}$$

This formulation proves sufficiently general to encompass all classical theories of interest: states of a classical system with $N$ degrees of freedom are represented by points in a $2N$-dimensional manifold $M$, which carries a symplectic form $\omega_{ab}$ (a closed, non-degenerate 2-form on $M$). The Hamiltonian is a function on $M$ which induces a Hamiltonian vector field $h^a$ via

$$h^a = \omega^{ab}\nabla_b H \tag{2.1.3}$$

where $\omega^{ab}$ is the numerical inverse of $\omega_{ab}$, that is, $\omega^{ac}\omega_{cb} = \delta^a_b$. The allowed motions on $M$ correspond to integral curves of $h^a$. Our notion of a classical theory is one in which all these elements are present. The combination $(M,\omega)$ is called a symplectic manifold. Therefore, a classical theory is a symplectic manifold together with a distinguished observable $H$ which determines the dynamics.

Observables are real-valued analytic functions on $M$, so the set $\mathfrak{A}$ of observables is a ring with the usual pointwise multiplication in $\mathfrak{A}$. However, the symplectic form $\omega$ can be used to induce a further algebraic structure on $\mathfrak{A}$ which has a deeper meaning. Define the Poisson bracket of observables $f$, $g \in \mathfrak{A}$ by

$$[f,g]_{\mathrm{P}} = \omega_{ab}\frac{\partial f}{\partial y_a}\frac{\partial g}{\partial y_b}. \tag{2.1.4}$$

This is antisymmetric between $f$ and $g$. In terms of the Poisson bracket, the dynamical equation (2.1.1) becomes

$$\frac{\mathrm{d}y_a}{\mathrm{d}t} = -[H, y_a]_{\mathrm{P}}. \tag{2.1.5}$$

The Poisson bracket turns $\mathfrak{A}$ into a Lie algebra.

There are various equivalent formulations of what we have just written which are useful in varying circumstances, depending on the problem at hand and the tools available to probe it. In order to pass from the familiar formulation, just in terms of configuration

space and canonical momenta, it is useful to view the configuration space variables $q_i$ as local coordinates on an $N$-dimensional configuration manifold $X$, and phase space as the cotangent bundle $T^*X$. The $(q^i, p_j)$ provide local coordinates on $T^*X$, and the symplectic form $\omega$ can be written

$$\omega = \mathrm{d}p_i \wedge \mathrm{d}q^i \tag{2.1.6}$$

Many physical systems admit a Lagrangian formulation in which the symplectic manifold $(M, \omega)$ can be derived from a simple expression. If the action $S$ is written as the integral of a one-form $L$, called the Lagrangian,

$$S = \int L, \tag{2.1.7}$$

then under deformations of the coordinates $q_i$, the change in the action will be

$$\delta S = \int e_i \delta q^i + \mathrm{d}\gamma, \tag{2.1.8}$$

where $\mathrm{d}\gamma[q^i, \delta q^j]$ is an exact form which gives a boundary term in the action. This term is conventionally discarded at the classical level, leaving only the field equations $e_i = 0$. The symplectic form is defined by $\omega = \delta\gamma$, with $\delta$ considered as the exterior differential on the space of deformations. For example, in the case of the Newtonian point particle one has

$$S = \int \frac{m}{2}\dot{q}^2 \,\mathrm{d}t, \quad \text{so} \quad \delta S = \int m\dot{q}\delta\dot{q} \,\mathrm{d}t = \int \mathrm{d}(-m\dot{q})\delta q + \int \mathrm{d}(m\dot{q}\delta q). \tag{2.1.9}$$

This leaves the familiar field equation $\ddot{q} = 0$, and the canonical sympectic form is

$$\omega = \delta(m\dot{q}\delta q) = m\delta\dot{q} \wedge \delta q = \delta p \wedge \delta q. \tag{2.1.10}$$

Not all classical theories admit a Lagrangian description. For example, a free chiral scalar field in two dimensions cannot be described in this way, but where it is available the Lagrangian formulation is usually the most computationally convenient (Deligne et al., 1999).

## 2.2. Quantum mechanics

This description of states as points in phase space and observables as functions of the state can be generalized in several interesting directions. Firstly, although the idea of a state as a point $(q^i, p_j)$ is quite sharp, it may happen for practical reasons that one cannot give such an exact characterization, or that one would like to deal with a certain number of (in principle distinct) states as equivalent. In either case, one is dealing with a state which

is not a point in $M$ but a region. The natural expression of this idea interprets a state as a measure on $M$: each state $\omega$ assigns to each observable $A$ a probability distribution $\omega_A(\lambda)$ on the real line. A state for which $\omega$ is an atomic measure is called pure. States where $\omega$ has extended support are called mixed.[1]

As a second generalization, one can notice that in the foregoing description we took the set of observables $\mathfrak{A}$ to be the ring of real-valued analytic functions on $M$ (Fadeev, 1999). This is a perfectly sensible choice, but it is stronger than necessary. Let $\mathfrak{A}$ be any vector space. The inner product[2]

$$\langle \omega | A \rangle = \int \lambda \, \mathrm{d}\omega_A(\lambda) \tag{2.2.1}$$

defines a duality between the set $\mathfrak{A}$ of operators and the set $\Omega$ of states. We assume that $\mathfrak{A}$ is complete in the sense of Appendix A: observables having the same inner product with all states are equal; therefore $\mathfrak{A}$ is separable. The observable $B$ is a function $f(A)$ of the observable $A$ if for all states $\omega$

$$\langle \omega | B \rangle = \int f(\lambda) \, \mathrm{d}\omega_A(\lambda). \tag{2.2.2}$$

This definition is not very easy to handle. Instead, one can make the more practical (but additional) technical assumption that observables $A$ lie in an algebra with a product $AB$ such that the notion of function (2.2.2) is compatible with the product.[3]

---

[1]The notion of pure and mixed states is usually only formally encountered in quantum mechanics where the definition can be quite mysterious, but there is no reason of principle why pure and mixed states should not also exist in classical mechanics. Indeed, the concept is more transparent when expressed in terms of the classical phase space. In fact, complete formulations of classical mechanics exist which are based on exactly this principle. An early example is the Koopman–von Neumann formalism, which is based on 'classical' wavefunctions defined on phase space (Mauro, 2003).

[2]It might appear more natural to take $\langle \omega | A \rangle = \int \omega_A(\lambda) \, \mathrm{d}\lambda$, which can be obtained by integrating by parts from this expression, but with this choice the definition of functional dependence in Eq. (2.2.2) is very much more natural. Nonetheless, some distinctions should carefully be borne in mind, of which the most important is that, despite appearances, $\mathrm{d}\omega_A(\lambda)$ is not a measure but the derivative of a measure.

[3]It is necessary to be quite careful about multiplication of observables. Although we are assuming that one can turn observables into an algebra, the multiplication is not quite algebraic but rather consists of a generalized sort of multiplication in which the product of operators $A(x)$, $B(x')$ at points $x$, $x'$ has an expansion in terms of other local operators $C_n(x)$ as a kind of Laurent series,

$$A(x)B(x') = \sum_{n=-\infty}^{\infty} C_n(x)(x - x')^n. \tag{2.2.3}$$

The useful feature of these generalizations is that the passage to quantum mechanics is now rather easy. One recovers classical mechanics immediately by supposing that $\mathfrak{A}$ is commutative, which was the choice we made above in the description of the classical world. In the case of quantum mechanics, $\mathfrak{A}$ is a complex associative algebra with involution (usually called complex conjugation in this context; such algebras are known as $*$-algebras). This is usually realized as an algebra of linear operators in some complex Hilbert space $H$, and the Lie bracket is defined in terms of the algebraic product as

$$[A, B]_{\mathrm{P}} = \frac{\mathrm{i}}{\hbar}(AB - BA). \tag{2.2.4}$$

The appearance of Planck's constant here follows on dimensional grounds (de Azcárraga and Izquierdo, 1995).

The simplest example is mechanics in one dimension. Here the phase space is $\mathbf{R}^2$ and points in phase space are pairs $(p, q)$. The Hilbert space is $L^2$. We replace the basic observables $p$ and $q$ by operators $\hat{p}$ and $\hat{q}$ acting on $L^2$, defined by

$$\hat{p} = -\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}q} \quad \text{and} \quad \hat{q} = q. \tag{2.2.5}$$

This is called the coordinate realization, or canonical quantization. For general observables $A$, one characterizes the action of $A$ by means of its kernel $A(q', q)$, viz.

$$(A\psi)(q') = \int_{-\infty}^{\infty} A(q', q)\psi(q)\,\mathrm{d}q \tag{2.2.6}$$

where $\psi$ is a state in $L^2$ (Deligne et al., 1999).

One must find such kernels by hand. An alternative scheme was introduced by Hermann Weyl in the early 1930s. For a general observable $A$ in $D$ dimensions, admitting a Fourier representation, one takes the Weyl symbol $\mathsf{W}(A)$ to be defined by

$$\mathsf{W}(A) = \int \widetilde{A(\xi)} \exp\left(\frac{\mathrm{i}}{\hbar}\xi^a \hat{y}_a\right) \mathrm{d}^{2D}\xi \tag{2.2.7}$$

where $y_a$ are coordinates on phase space. The function $\widetilde{f(\xi)}$ is the Fourier transform of $f(x)$, and the $\mathsf{W}(A)$ form an algebra, called the Weyl algebra, which quantizes the subset of

This is called the operator product expansion (OPE). In general, it shows that the product of two operators at a point is singular, although there are usually only a finite number of singular terms. The OPE is extremely useful in practice and is an indispensible tool in string theory, but it is unnecessary for the applications we are going to describe so we will omit it despite its great intrinsic interest. The OPE is related to generalized kind of algebra known as an affine Lie algebra or Kac–Moody algebra which appears in string theory as the Virasoro algebra of the energy–momentum tensor (Polchinski, 1998).

classical observables admitting a Fourier transform. Using the Baker–Campell–Hausdorff formula and supposing that the commutator of the $\hat{y}_a$ closes on a constant shows that the product $\mathsf{W}(A)\mathsf{W}(B)$ of two Weyl symbols gives

$$\mathsf{W}(A)\mathsf{W}(B) = \int \widetilde{A(\xi)}\widetilde{B(\eta)} \exp\left(\frac{\mathrm{i}}{\hbar}(\xi^a + \eta^a)\hat{y}_a - \frac{\mathrm{i}}{2\hbar}\xi^a\eta^b\omega_{ab}\right) \mathrm{d}^{2D}\xi\, \mathrm{d}^{2D}\eta. \tag{2.2.8}$$

By a suitable change of variables and some lengthy rearrangement, it can be shown that this is the same as the Weyl symbol $\mathsf{W}(A \star B)$ of an observable

$$(A \star B)(y) = \exp\left(\frac{\mathrm{i}\hbar}{2}\omega_{ab}\frac{\partial}{\partial y_a}\frac{\partial}{\partial y'_b}\right) A(y)B(y')\Big|_{y'\to y} = 1 + \frac{\mathrm{i}\hbar}{2}[A, B]_{\mathrm{P}} + O(\hbar^2). \tag{2.2.9}$$

The product $\star$ is called the Moyal product, and was introduced in 1949 by Moyal (then a graduate student) to study quantum statistical mechanics; it was first called the Moyal product by Groenewald. The Moyal commutator is

$$\frac{\mathrm{i}}{\hbar}(A \star B - B \star A) = [A, B]_{\mathrm{P}} + O(\hbar). \tag{2.2.10}$$

Therefore both classical and quantum mechanics can be realized in terms of the same objects, functions on phase space, but with the structure constants of the algebraic operations (the product and the Lie bracket) in quantum mechanics being defined as a power series in positive powers of $\hbar$, with the zero-order term coinciding with the structure constants of classical mechanics. In this sense, quantum mechanics is a strict deformation of classical mechanics, with the Planck constant $\hbar$ as the corresponding deformation parameter. The precise details of the deformation are controlled, with perfect inevitability, by the classical symplectic form $\omega_{ab}$. This is one of the most concise and beautiful explanations of the relationship between quantum and classical mechanics. In fact, more is true. It can be shown that classical mechanics is unstable and that quantum mechanics is the essentially unique stable deformation of classical mechanics into a non-equivalent stable structure (Dito and Sternheimer, 2002; Sternheimer, 1998; Weinstein, 1994)

To relate the Weyl symbol $\mathsf{W}(A)$ of an observable $A$ to its kernel, one lets the Weyl symbol operate on a state $\psi(q)$, which, after integrating by parts and discarding a surface term, can be written

$$\mathsf{W}(A)\psi(q) = \int_{-\infty}^{\infty} \mathrm{d}\xi \int_{-\infty}^{\infty} \mathrm{d}\eta\, \widetilde{A(\xi,\eta)}\, \mathrm{e}^{-\mathrm{i}\hbar\xi q}\psi(\mathrm{i}\hbar\frac{\partial}{\partial q})\mathrm{e}^{-\mathrm{i}\eta q/\hbar}. \tag{2.2.11}$$

On the other hand the same quantity in terms of the integral kernel is

$$(A\psi)(q) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} \mathrm{d}q' \int_{-\infty}^{\infty} \mathrm{d}p\, \mathrm{e}^{\mathrm{i}pq'/\hbar}\psi(\mathrm{i}\hbar\frac{\partial}{\partial p})\left[f(p, \frac{q+q'}{2})\mathrm{e}^{-ipq/\hbar}\right], \tag{2.2.12}$$

where we have again integrated by parts and discarded a boundary contribution, and the integral kernel for $A$ is defined by (Fadeev, 1999)

$$A(q,q') = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} \mathrm{d}p\, f(p, \frac{q+q'}{2})\mathrm{e}^{-\mathrm{i}p(q-q')/\hbar}. \tag{2.2.13}$$

Clearly (2.2.11) can be obtained from (2.2.12) by suitable relabelling of variables. The quantity $f$ is confusingly also known as the symbol of $A$, in this case without the attendant qualifier Weyl. As a special case we write the symbol (in this sense) of the Hamiltonian $H$ as $h(p,q)$.

## 2.3. Quantum field theory

The discussion so far has described paths on phase space as functions of a single parameter taking values in $\mathbf{R}$. By convention, this parameter is associated with time although this does not have to be the case. To promote the discussion to the level of quantum field theory, one replaces $\mathbf{R}$ by Minkowski space (in four dimensions) $M_4$ with metric diag$(-,+,+,+)$. The analogue of the coordinates $q^i(t)$ are scalar fields $\phi^i(x^a)$ which vary over $M_4$.

There are several important features of quantum field theory which it is necessary to outline here, because we shall make much use of them later. The first, outlined in Section 2.3.1, is the use of the path integral to calculate transition amplitudes. Almost all of the later work we shall describe relies entirely on this technology. The second (Section 2.3.2) is the expansion of the theory into diagrams, although in this work we will not place too much emphasis on diagrammatica since it is not especially suited to many of the later applications. (However, the calculation of radiative corrections to the quintessence mass described in Chapter 6 does depend on this technique.) We also outline the basics of perturbative renormalization. This is discussed in Section 2.3.3, including in some detail the calculation of radiative corrections to the propagator, which is most important in applications. In particular, this provides an opportunity to introduce in a straightforward context some of the mathematical methods which will prove important later. The theory of radiative corrections to the propagator also has some relevance to the theory of early

universe inflation, which will be set out in Section 5.7 and Section 7.2; we postpone discussion of this matter until then. Finally we also need to compute the quantum effective potential (Section 2.3.4).

**2.3.1. Path integral transition amplitudes.** Consider the dynamical equation (2.1.1) for some observable $A$ with boundary condition $A(t_0) = A_0$. The solution is

$$A(t) = U^{-1}(t) A_0 U(t) \tag{2.3.1}$$

where the time evolution operator $U(t)$ is defined by

$$U(t) = \exp\left(-\frac{\mathrm{i}}{\hbar} H(t - t_0)\right), \tag{2.3.2}$$

as can be proved at once by simple differentiation. We wish to find the kernel corresponding to $U(t)$. Because the Weyl quantization is not a homomorphism, which would behave nicely with respect to multiplication of operators, one cannot simply substitute the kernel of the Hamiltonian into the power series for the exponential to obtain the kernel for $U(t)$. The solution is simply to deal with infinitesimal time slices $\Delta = t - t_0$, where products of operators do not occur and $U(t)$ is just

$$U(t) \simeq 1 - \frac{\mathrm{i}}{\hbar} H\Delta. \tag{2.3.3}$$

If we divide the macroscopic time interval $t - t_0$ into $N$ small slices, each of size $\Delta$, then the full time evolution operator can be recovered just by repeated composition,

$$\exp\left(-\frac{\mathrm{i}}{\hbar} H(t - t_0)\right) = \left(1 - \frac{\mathrm{i}}{\hbar} H\Delta\right)^N. \tag{2.3.4}$$

The infinitesimal kernel of $U(t)$, making use of the correspondence (2.2.13) between the symbol, Weyl symbol and integral kernel, is therefore

$$\begin{aligned} U(q, q_0; \Delta) &\simeq \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{d}p\, \mathrm{e}^{\mathrm{i}p(q-q_0)/\hbar} \left(1 - \mathrm{i}h(p, \frac{q+q_0}{2})\Delta\right) \\ &= \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{d}p\, \mathrm{e}^{\mathrm{i}\hbar^{-1}\left[p(q-q_0) - h(p, \frac{q+q_0}{2})\Delta\right]} + O(\Delta^2) \end{aligned} \tag{2.3.5}$$

Composing this $N$ times gives the result we shall require,

$$\begin{aligned} U(q, q_0; t, t_0) = \int \cdots \int \mathrm{e}^{\mathrm{i}\hbar^{-1}[p_N(q-q_{N-1}) + \cdots + p_1(q_1 - q_0)]} \times \\ \mathrm{e}^{-\mathrm{i}\hbar^{-1}\left[h(p_N, \frac{q_N+q_{N-1}}{2})\Delta - \cdots - h(p_1, \frac{q_1+q_0}{2})\Delta\right]} \frac{\mathrm{d}q_1 \mathrm{d}p_1}{2\pi\hbar} \cdots \frac{\mathrm{d}q_{N-1} p_{N-1}}{2\pi\hbar} \frac{\mathrm{d}p_N}{2\pi}. \end{aligned} \tag{2.3.6}$$

As one sends $N$ to $\infty$, the term in the exponent goes over to

$$\exp\left(\mathrm{i}\hbar^{-1}\int_{t_0}^{t}[p\dot{q}-h]\,\mathrm{d}t'\right). \tag{2.3.7}$$

This is essentially the action integral, since $h$ is related to the Hamiltonian. The different choices for generating $h$ from the classical Hamiltonian $H$ correspond to different ways of interpreting the integral and its associated measure,

$$\prod_t \frac{\mathrm{d}q(t)\mathrm{d}p(t)}{2\pi\hbar} = [\mathrm{d}p][\mathrm{d}q] \tag{2.3.8}$$

which is the Liouville measure on paths in phase space. Therefore the time evolution operator is given by a path integral over trajectories in phase space. This explains the central importance in quantum mechanics of methods in functional analysis, and the functional integral in particular. It is important to note carefully, however, that the quantity appearing under the integral here is not exactly the Lagrangian, which is $p\dot{q}-h$ where $p$ and $q$ are related via Hamilton's equations. No such requirement is in force here. Indeed, the $p$ and $q$ may vary over trajectories in phase space without constraint.

A real polarization of a symplectic manifold, such as $(M,\omega)$, is a splitting of the coordinates into positions and momenta. Such a splitting is rarely unique. In the present case, the boundary condition on the trajectories in phase space is that $q(t_0)=q_0$ and $q(t)$ coincides with the other end of the path, whereas the $p$ are unconstrained at either end of the trajectory.

If the Hamiltonian function $h(p,q)$ is quadratic in $p$ then the $[\mathrm{d}p]$ integral can be done exactly, leaving (up to constant factors which we have absorbed into the normalization)

$$U(q,q_0;t,t_0)=\int[\mathrm{d}q]\,\exp\left(\mathrm{i}\int_{t_0}^{t}L[q,q_0]\right) \tag{2.3.9}$$

where $L$ is the honest Lagrangian form which depends on the boundary conditions $q$, $q_0$.

**2.3.2. The expansion into diagrams.** The description of quantum mechanics in terms of functional integrals, which we have derived here from Weyl's quantization of the phase space $(M,\omega)$ although there are alternative routes, provides the most compact, practical approach to quantum mechanics. Among the most important reasons for preferring the path integral is the ease with which topologically non-trivial configurations can be handled, because of the direct description of the integral in terms of paths on phase space,

or configuration space in the special case that the Hamiltonian is quadratic in the momenta. The technology we developed in Appendix A allows us to make some progress in the direct evaluation of paths integrals, and therefore provides an entry to direct, concrete calculations in quantum mechanics.

If one could calculate all path integrals exactly we would be finished, because it would be possible to answer essentially any question one might wish to ask. This is intuitively easy to see, since the integral (2.3.9) supplies an explicit form for the time evolution equation, which is enough to settle any question of dynamics. To see this in detail, observe that one uses $U(q, q_0; t, t_0)$ to evolve a state $\psi_{t_0}(q)$ at time $t_0$ into a state at time $t$,

$$\psi_t(q) = \int_{-\infty}^{\infty} \mathrm{d}q_0 \, U(q, q_0; t, t_0)\psi_{t_0}(q_0). \tag{2.3.10}$$

This result can be iterated. To obtain the time evolution operator between two times $t_a$ and $t_b$, in terms of time evolution between $t_a$ and $t_c$, and $t_c$ and $t_b$, where $t_a < t_c < t_b$, one has

$$U(q_b, q_a; t_b, t_a) = \int_{-\infty}^{\infty} \mathrm{d}q_c \, U(q_b, q_c; t_b, t_c)U(q_c, q_a; t_c, t_a) \tag{2.3.11}$$

This is true provided the action appearing in $U$ is additive over neighbouring regions, that is,

$$S[t_a, t_b] = S[t_a, t_c] + S[t_c, t_b]. \tag{2.3.12}$$

This is trivially true for many simple actions, but care must be taken in the case of gravity where an extra boundary contribution (the Gibbons–Hawking term) must be added to ensure that additivity of the action holds. We will discuss this subtlety later, when discussing quantum cosmology in Chapter 8.

Unfortunately, it is largely impossible to evaluate any path integrals directly. Indeed, only in the very special case of a Gaussian measure (which we discussed in Appendix A) can one perform integrals exactly, and even in this case the path integral suffers from a potential ill-definedness owing to difficulties in correctly intepreting the measure. A Gaussian measure on phase space is equivalent to a quadratic Lagrangian, of the form $\mathcal{L} = \phi\triangle\phi$ for a field $\phi$ and a negative-definite, self-adjoint operator $\triangle$. In the sequel, it is convenient to factor a minus sign out of $\triangle$, so that $\mathcal{L}$ is manifestly positive definite and enters into the measure with an explicit minus:

$$U = \int [\mathrm{d}\phi] \, \exp\left(-\frac{\mathrm{i}}{\hbar}\int_M \mathrm{d}v \, \phi\triangle\phi\right) = \left(2\pi\hbar\mu^2\right)^{-\zeta(0)/2} \exp\left(\frac{1}{2}\zeta'(0)\right) \tag{2.3.13}$$

where $M$ is a causal region bounded by spacelike hypersurfaces of some manifold with Lebesgue measure $\mathrm{d}v$; $\zeta(s)$ is the $\zeta$-function of $\mathrm{i}\triangle$; and $\mu$ is a constant with the dimensions of mass which is present on dimensional grounds. This is just the result of Eq. (A.2.11). Such a theory is called free. It can be evaluated exactly in terms of the $\zeta$-function of the quadratic (or free) kernel $\triangle$. In addition (Section A.2.1) one can perform these integrals with the possible insertion of an arbitrary polynomial $P(\phi)$ of $\phi$ in the integral, as in $\int[\mathrm{d}\phi]\ P(\phi)\mathrm{e}^{-\mathrm{i}\hbar^{-1}S}$.

The technique of Feynman graphs allows one to write down an asymptotic series for the path integral transition amplitude of an arbitrary quantum field theory in the neighbourhood of a free theory. Consider some general theory for $\phi$. This will contain a free field part plus a collection of higher order terms,

$$\mathcal{L} = \phi\triangle\phi + \sum_{n=0}^{\infty} \frac{g_n}{n!}\phi^n, \tag{2.3.14}$$

where the $g_n$ are finite positive or negative coupling constants, any number of which may be zero. This is to be understood as a deformation of the original Lagrangian in the infinite-dimensional space of possible theories described by the $g_n$. To evaluate this, we (formally) write

$$Z = \int[\mathrm{d}\phi]\ \exp\left(-\frac{\mathrm{i}}{\hbar}\int_M \mathrm{d}v\ \phi\triangle\phi\right) \sum_{m=0}^{\infty} \frac{1}{m!}\left(\int_M \mathrm{d}v \sum_{n=0}^{\infty} \frac{-\mathrm{i}g_n}{\hbar n!}\phi_n\right)^m. \tag{2.3.15}$$

We have written the transition amplitude as $Z$ rather than $U$ to indicate that we are now enforcing a particular set of boundary conditions: namely, that $Z$ gives the amplitude for transitions between the quantum vacuum state. This is the usual choice in quantum field theory, and we adopt it exclusively henceforth. Because of its definition, $Z$ is usually called the vacuum–vacuum amplitude or vacuum persistance amplitude. This amplitude is a power series in the the $g_n$ whose terms are path integrals of polynomial functions of $\phi$ against the free field measure. As a result, we can use the results of Section A.2.1 to calculate the individual terms in the power series. This is called the perturbation series for the theory, or sometimes the Feynman series. There is an important interpretation of each term, due to Feynman. In order to simplify the discussion, we assume that only the lowest-order deformation in (2.3.14) is present, which has the form $g_3\phi^3/6$.[4]

[4]This is usually called the $\phi^3$ theory. As a matter of fact it is physically rather unsatisfactory, since the total energy integral is unbounded below. The attraction of the $\phi^3$ theory lies in the simplicity of its

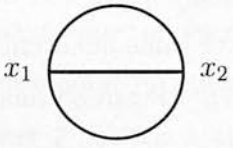
Consider the term of order $m$ in the Feynman series. This has the form

$$Z_m = \int [\mathrm{d}\phi]\, \frac{1}{m!} \int \cdots \int \frac{-\mathrm{i}g_3}{6\hbar} \cdots \frac{-\mathrm{i}g_3}{6\hbar}\, \mathrm{d}v_1 \cdots \mathrm{d}v_m\, \phi^3(x_1) \cdots \phi^3(x_m)\, \exp\left(-\frac{\mathrm{i}}{\hbar} \int_M \mathrm{d}v\, \phi \triangle \phi\right). \tag{2.3.16}$$

From the general results obtained in Section A.2.1, in particular Wick's theorem (A.2.8), this is the same as

$$Z_{2m} = \frac{1}{(2m)!} \left(\det \frac{\mathrm{i}\triangle}{2\pi\hbar\mu^2}\right)^{-1/2} \left(\frac{-\mathrm{i}g_3}{6\hbar}\right)^{2m} \int \cdots \int \mathrm{d}v_1 \cdots \mathrm{d}v_{2m} \sum_{\text{pairings}} \prod_{(x,y))} (-\mathrm{i}\hbar)\triangle^{-1}(x,y), \tag{2.3.17}$$

which is non-zero only for even terms. As an equivalent formulation of the same expression, we can say that this is equivalent to a graph with $2m$ 3-valent vertices labelled with a spacetime coordinate $x_i$, with the vertices all joined by arcs. At each vertex, one places the coupling constant $-\mathrm{i}g_3/6$ and contracts along the arcs using the quadratic form $-\mathrm{i}\hbar\triangle^{-1}$. Finally, one integrates over the spacetime position of each vertex, and divides by the symmetry factor $1/(2m)!$ which in this case counts the automorphisms of the graph into itself by exchange of vertices. For example, the lowest order term $Z_2$ corresponds to the diagram




Such diagrams are called Feynman diagrams, and the corresponding integrals are called Feynman integrals. The rules for obtaining Feynman integrals from Feynman diagrams are called the Feynman rules. In this case, the Feynman integral is

$$-\frac{\mathrm{i}\,3!}{2}\frac{g_3^2\hbar}{36} \int \mathrm{d}v_1 \int \mathrm{d}v_2\, \triangle^{-1}(x_1,x_2)\triangle^{-1}(x_1,x_2)\triangle^{-1}(x_1,x_2). \tag{2.3.18}$$

The multiplicity 3! arises from the number of possible pairings of the three arcs leaving the left-hand vertex with arcs entering the right-hand vertex. This interpretation of the coefficients of $Z$ is called the expansion of the theory into diagrams. The same interpretation is possible with a more general deformation of the free theory Lagrangian, including more interaction types. For example, in a Lagrangian with both $\phi^3$ and $\phi^4$ interactions present,

$$\mathcal{L} = \frac{1}{2}\phi\triangle\phi + \frac{g_3}{3!}\phi^3 + \frac{g_4}{4!}\phi^4 \tag{2.3.19}$$

---

diagram expansions. The same ideas and methods that we will outline in the context of the $\phi^3$ theory apply equally to more complicated, realistic theories.

one includes at any given order $N$ diagrams which comprise any number $M \leqslant N$ of 4-valent vertices and $N - M$ 3-valent vertices, together with arcs joining each vertex. One places $-\mathrm{i}g_3/3!\hbar$ at each 3-valent vertex, $-\mathrm{i}g_4/4!\hbar$ at each 4-valent vertex, contracts vertices with $-\mathrm{i}\hbar\triangle^{-1}$, and integrates over the positions of all vertices to obtain the corresponding Feynman integral.

There is no reason why the Feynman series should converge. Although not much is understood at a precise level about its convergence properties, Dyson noticed soon after its inception that the number of diagrams at order $n$ grows like $n!$, so one expects that the Feynman series should have radius of convergence zero. For this reason one should strictly interpret it as an asymptotic series in the coupling constants: there is no guarantee that at each order the contribution from Feynman diagrams decreases, but this does happen in the few low-order cases which have so far been studied in detail. Indeed, almost all precision particle physics experiments are carried out using the perturbative expansion.

So far we have been dealing only with transition amplitudes between states corresponding to the quantum vacuum at a time $t$. Integrals with the insertion of a non-trivial polynomial have a special meaning, which can be related to more general states in the quantum theory. Suppose we have some quantum field theory with action $S[\phi]$, which we couple to a classical current $J$. The transition function in the presence of these currents is

$$Z[J] = \int [\mathrm{d}\phi]\ \exp\left(-\frac{\mathrm{i}}{\hbar}S[\phi] + \frac{\mathrm{i}}{\hbar}\int_M \mathrm{d}v\ \phi J\right) \tag{2.3.20}$$

The functional $Z[J]$ is sometimes called the partition function for the theory. The rules for expanding $Z[J]$ into diagrams are the same as before, except now the theory contains a new kind of vertex to which a single $\phi$ line is attached with position-dependent coupling factor $-\mathrm{i}\hbar^{-1}J(x)$. The coefficient of $(-\mathrm{i}\hbar^{-1}J)^n$ in a series expansion of $Z[J]$ in powers of $J$ is the sum of all diagrams which contain $n$ $J$-vertices with the coupling factor at each vertex stripped off, leaving only one leg of the Green's function $\triangle^{-1}$ exposed. In this context, one usually calls $-\mathrm{i}\hbar\triangle^{-1}$ the propagator, and refers to its exposed leg as an external line. To find these coefficients, one differentiates with respect to $J(x)$ and then sets $J = 0$. For example, one can obtain the sum of all diagrams with two external lines as

$$\frac{1}{2}\left(-\mathrm{i}\hbar\frac{\delta}{\delta J(x_1)}\right)\left(-\mathrm{i}\hbar\frac{\delta}{\delta J(x_2)}\right)Z[J]_{J=0}. \tag{2.3.21}$$

This is just the definition of what we mean by a functional Taylor series. However, one can do these derivatives exactly. Each derivative pulls a factor $\phi$ down into the integrand,

leaving

$$\frac{1}{n!}\underbrace{\left(-\mathrm{i}\hbar\frac{\delta}{\delta J(x_1)}\right)\cdots\left(-\mathrm{i}\hbar\frac{\delta}{\delta J(x_n)}\right)}_{n\text{ times}} Z[J]_{J=0} = \int [\mathrm{d}\phi]\ \phi(x_1)\cdots\phi(x_n)\ \exp\left(-\frac{\mathrm{i}}{\hbar}S[\phi]\right). \tag{2.3.22}$$

Therefore, path integrals containing arbitrary monomials in the integrand correspond to Feynman diagrams containing a given number of external legs, one leg for each factor of $\phi$ in the monomial. One calls these functions the correlation functions of the theory, or $n$-point functions where it is necessary to be specific, and writes them as

$$\langle T[\phi(x_1)\phi(x_n)]\rangle = \int [\mathrm{d}\phi]\ \phi(x_1)\cdots\phi(x_n)\ \exp\left(-\frac{\mathrm{i}}{\hbar}S[\phi]\right), \tag{2.3.23}$$

where $T$ is a time-ordering function that rewrites its argument in order of decreasing time, from left to right. In later equations, the presence of this time-ordering is understood. For example, the 2-point function $\langle\phi(x_1)\phi(x_2)\rangle$ of the $\phi^3$ theory includes the diagram


It is by this means that we are able to introduce states other than the quantum vacuum into our calculations. For example, the two-point function corresponds to a process in which a 1-particle state evolves into another 1-particle state. There is no reason of principle why one should not consider processes in which an $m$-particle in state evolves into an $n$-particle out state. Feynman's diagrams provide a concrete interpretation of the processes by which in-states evolve into out-states: particles move in spacetime, as described by the propagators. At vertices incoming particles are destroyed and outgoing particles are created.

This description is adequate for the theory of a real scalar field, which we have been considering until now. However, for more sophisticated examples such as a complex scalar field or a Dirac fermion, one should notice that a quantum field such as $\phi(x)$ has the interpretation of creating a particle at spacetime position $x$, whereas the adjoint $\phi^\dagger(x)$ destroys such a particle. For this reason (using, for example, time-reversal invariance), out-state particles should appear as an adjoint in the path integral.

**2.3.3. Perturbative renormalization.** Unfortunately, many Feynman integrals are divergent. This is not the same as any overall divergence of the Feynman series, which might

only be an artefact of performing the Taylor expansion: in this case, it is an individual coefficient in the expansion which is divergent. Such bad behaviour would preclude any attempt to deal with quantum field theory using the Feynman expansion.

The problem of infinities has plagued quantum field theory from its very early days. Of course, at one level infinities are also present in the classical world, appearing for example as core divergences in the Newton or Coulomb potentials, or in the infinite classical self-energy of an isolated charge such as an electron. These have the form, respectively,

$$V_N = -\frac{Gm_1m_2}{r}, \quad V_C = -\frac{q_1q_2}{4\pi\varepsilon_0 r^2} \quad \text{and} \quad m_e = \frac{e^2}{6\pi c^2 r} \tag{2.3.24}$$

and all diverge as the radial coordinate $r$ approaches zero. Nonetheless, the presence of these classical divergences is disturbing, and not a justification for the presence of divergences in the quantum theory: over the last century, the quantum programme has consisted more or less in the systematic removal of infinites from classical mechanics, and it would be disturbing to see these successes disrupted by the appearence of new, less controllable infinites in the perturbation series. In fact, such a clamity does not occur, but to see this in detail requires the use of renormalization techniques. Indeed, to date, almost all classical infinites have been swept away when given a more detailed quantum mechanical treatment, with the single notable exception of infinities which correspond to singularities in general relativity. Although conjectural ideas exist, no compelling solution is yet available to handle such gravitational singularities.[5]

One particularly useful example is the two-point function of the $\phi^3$ theory. This calculation provides tools and notation for the very similar calculation of radiative corrections to the quintessence propagator which will be encountered in Section 6.7. In terms of path integral expressions as derived above, the two-point function satisfies

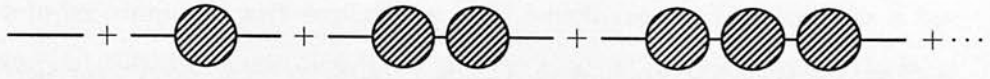
$$\langle\phi(x_1)\phi(x_2)\rangle = \int [\mathrm{d}\phi]\ \phi(x_1)\phi(x_2)\exp\left[-\frac{\mathrm{i}}{\hbar}\int_M \mathrm{d}v\ \left(\frac{1}{2}\phi\triangle\phi + \frac{g}{6}\phi^3\right)\right]. \tag{2.3.25}$$

[5]Nonetheless, progress is being made. For example, in the AdS/CFT correspondence (to be studied later) black hole states in a gravitational theory correspond to thermal states in a conformal field theory with one dimension less, where the equivalence is supposed to hold with all quantum corrections included. The thermal state is non-singular, so although we cannot yet compute the details of quantum corrections to the gravitational theory, when such corrections are taken into account they must presumably smooth the divergence away. Exactly how the details of this smoothing are worked out, and what the precise effect in the gravitational world looks like, are not yet accessible to investigation, but there is some hope that this could change in the not too distant future.

The expansion of this integral into diagrams is the sum of all possible diagrams with two external vertices. This includes both connected and disconnected graphs. Since there must be two external vertices, it is easy to see that the sum of all disconnected graphs must take the form


where each external leg is connected to the sum of all vacuum diagrams (here denoted by the shaded circles); in other words, this is the square of the one-point function $\langle\phi(x)\rangle$. In almost all cases, this must vanish. For fields of higher rank than a scalar, this happens just by Lorentz invariance and for scalar fields it is easy to see from (2.3.22) that if $\langle\phi(x)\rangle$ is not zero, then we are not expanding around an extremum of the action. In this case, the vacuum of the theory is not stable and we should expect strange effects to occur. We will always assume that the one-point function vanishes, in which case disconnected graphs can be ignored.

On the basis of these assumptions, it is possible to restrict attention to connected Feynman diagrams. One defines a 1-particle irreducible graph to be any diagram which cannot be disconnected by cutting through only one internal line, and the sum of all 1-particle irreducible graphs to be $\mathrm{i}\hbar^{-1}\Sigma$, where the initial and final propagators are stripped off. The connected contribution to the two-point function then has to take the form of a sum over all chains of 1-particle irreducible diagrams, with propagators added correctly,

$\langle\phi(x_1)\phi(x_2)\rangle =$  $+\cdots$

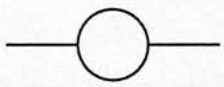
(2.3.26)

According to the standard rules for handling Feynman diagrams which we outlined above, this sum evaluates to

$$\begin{aligned}\langle\phi(x_1)\phi(x_2)\rangle &= -\mathrm{i}\hbar\triangle^{-1} + (-\mathrm{i}\hbar\triangle^{-1})(\Sigma\triangle^{-1}) + (-\mathrm{i}\hbar\triangle^{-1})(\Sigma\triangle^{-1})(\Sigma\triangle^{-1}) \\ &\quad + (-\mathrm{i}\hbar\triangle^{-1})(\Sigma\triangle^{-1})(\Sigma\triangle^{-1})(\Sigma\triangle^{-1}) + \cdots \\ &= -\mathrm{i}\hbar\triangle^{-1}\sum_{n=0}^{\infty}(\Sigma\triangle^{-1})^n = \frac{-\mathrm{i}\hbar}{\triangle - \Sigma}.\end{aligned} \tag{2.3.27}$$

It was in order to obtain a conveniently handled series such as this that we chose to strip propagators away from $i\hbar^{-1}\Sigma$. If the theory were free, then the two-point function would just equal $-i\hbar\triangle^{-1}$, which is therefore called the bare propagator, whereas the result for $\langle\phi(x_1)\phi(x_2)\rangle$ given above is known as the dressed propagator: it is the two-point function with all radiative corrections included. This is an example of how one can use the Feynman series to evaluate functional integrals for quite general quantum field theories — provided, of course, that the theory is in the neighbourhood of a free theory. The function $i\hbar^{-1}\Sigma$ is often called the self-energy of the particle: it is this function which was found to be divergent in the early days of quantum field theory.

Divergences can be found at the first non-trivial order in perturbation theory. It is conventional to write the $g$-expansion of $\Sigma$ as $\Sigma = \sum_n g_n\Sigma_n$, in which case the lowest order contribution is $\Sigma_2$, which has the form

$i\tilde{\Sigma}_2 =$ 

(2.3.28)

where we write $\tilde{\Sigma}_2$ for the present in order to indicate that initial and final propagators have not yet been stripped off. We have done this so that the transition to momentum space, worked out below, is more transparent. Working in Minkowski space, the corresponding Feynman integral which translates this graph into an analytical function satisfies

$$i\hbar^{-1}\tilde{\Sigma}_2(x_1,x_2) = \int \mathrm{d}^4v_1\mathrm{d}^4v_2 \left(\frac{-ig_3}{6}\right)^2 (-i\hbar)^4\triangle^{-1}(x_1,v_1)\triangle^{-1}(v_1,v_2)\triangle^{-1}(v_2,v_1)\triangle^{-1}(v_2,x_2) \tag{2.3.29}$$

In cases where the configuration space is simple,[6] it is often useful to switch to Fourier modes. If we choose the free kernel around which we are deforming to correspond to a

[6]If this is not the case, then it can sometimes be convenient to leave the calculation with the Feynman rules in their configuration space formulation. In fact, this is exactly what we will do when deriving the Feynman rules on a brane compactification, where the "coupling constants" in fact involve $\delta$-functions in order to restrict their support to the worldvolume of the brane.

particle of mass $m$, then (with our choice of spacetime signature[7])

$$\triangle(x_1, x_2) = (-\partial_{x_1}^2 + m^2)\delta_{\mathrm{D}}(x_1 - x_2) \quad \text{so} \quad \triangle(x_1, x_2) = \int \frac{\mathrm{d}^4 k}{(2\pi)^4}(k^2 + m^2)\mathrm{e}^{\mathrm{i}k\cdot(x_1 - x_2)}. \tag{2.3.30}$$

The dressed propagator, or full two-point function, then corresponds to the expression (compare (2.3.27))

$$\langle \phi(x_1)\phi(x_2)\rangle = -\mathrm{i}\hbar \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{k^2 + m^2 - \Sigma(k)} \mathrm{e}^{\mathrm{i}k\cdot(x_1 - x_2)} \tag{2.3.31}$$

where $\int \mathrm{d}^4 k (2\pi)^{-4}\, \Sigma(k)\mathrm{e}^{-\mathrm{i}k\cdot x}$ is the Fourier transform of the self-energy function $\Sigma$. In this representation, it is easy to see that the effect of the self-energy is to introduce a $k$-dependent modification of the particle mass. For this reason, when we calculate a final expression for the mass of the particle in this theory, we do not expect it to coincide with $m$. The deformation of the Lagrangian takes the theory away from the point corresponding to the free kernel, and as it does so it takes all other parameters in the theory such as masses and coupling constants with it. The real particle mass is called the renormalized mass, and it must be chosen to match with experimental data returned from particle accelerators and similar experiments. There is no such requirement imposed on the parameter $m$, which is called the bare mass. This is the content of renormalization: it has nothing to do with cancellation of divergences in itself. It is merely a method for describing which numbers in the theory are to be compared with experiment, and which are not physical. As a by-product of working out this division in detail, we will see that all divergences are, in fact, cancelled.

[7]It is easy to decide which sign the mass of the particle should enter with, by remembering that the appropriate relativistic invariant is just the square of the particle momentum, $p^2$. With our choice of spacetime signature, diag$(-,+,+,+)$ this is given by minus the mass, so upon Fourier transforming one obtains $-p^2 - m^2$ which is zero, correctly, when the particle is on the mass shell. This happens because the purpose of the field equation is to restrict solutions at the classical level to have support only on the mass shell; away from the mass shell, the field equation $\triangle\phi = 0$ forces $\phi = 0$.

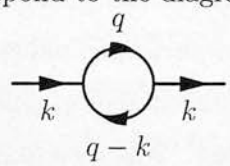
Carrying out the replacement of all bare propagators by (2.3.30) and doing the coordinate space integrals to give two $\delta$-functions gives

$$\mathrm{i}\hbar^{-1}\Sigma_2 = \left(\frac{-\mathrm{i}g_3}{6}\right)^2 \int \frac{\mathrm{d}^4k_1}{(2\pi)^4}\frac{\mathrm{d}^4k_2}{(2\pi)^4}\frac{\mathrm{d}^4k_3}{(2\pi)^4}\frac{\mathrm{d}^4k_4}{(2\pi)^4}\frac{\mathrm{i}\hbar}{k_1^2+m^2}\frac{\mathrm{i}\hbar}{k_2^2+m^2}\frac{\mathrm{i}\hbar}{k_3^2+m^2}\frac{\mathrm{i}\hbar}{k_4^2+m^2}\times$$
$$(2\pi)^4\delta_{\mathrm{D}}(-k_1+k_2-k_3)(2\pi)^4\delta_{\mathrm{D}}(-k_2+k_3+k_4)\mathrm{e}^{\mathrm{i}k_1\cdot x_1}\mathrm{e}^{\mathrm{i}k_4\cdot x_2} \tag{2.3.32}$$

This is equivalent to a reformulation of the Feynman rules we outlined above in which one places a momentum $k$ on each arc of the diagram. It is then usual to attach an arrow to the arc in order to give a sense to this momentum, although both these sense of the momentum and the arrows themselves are mere matters of convention. At each vertex, one places the coupling constant $-\mathrm{i}g$ together with a momentum-conservation term $(2\pi)^4\delta(\sum k)$, where by $\sum k$ we mean the sum of the momenta flowing into the vertex. For each arc, one includes a propagator factor

$$-\frac{\mathrm{i}\hbar}{(2\pi)^4}\frac{1}{k^2+m^2}; \tag{2.3.33}$$

and one adds terms $\mathrm{e}^{\mathrm{i}k\cdot x}$ for each ingoing particle of momentum $k$ at position $x$, and $\mathrm{e}^{-\mathrm{i}k\cdot x}$ for each outgoing particle of momentum $k$. To obtain the final Feynman amplitude, one integrates over all momenta. These are called the momentum space Feynman rules. In the present case, the rules for $\Sigma_2$ correspond to the diagram




where we have carried out the $\delta$-function integrals. If we now strip away the initial and final propagator factors, and the ingoing and outgoing momentum factors, we are left with an expression for $\Sigma_2(k)$

$$\mathrm{i}\hbar^{-1}\Sigma_2(k) = \frac{g_3^2}{36}\int \mathrm{d}^4q\,\frac{1}{q^2+m^2}\frac{1}{(q-k)^2+m^2} \tag{2.3.34}$$

To evaluate this, one makes use of Feynman's integral formula,

$$\frac{1}{AB} = \int_0^1 \mathrm{d}x\,\frac{1}{(xA+(1-x)B)^2}. \tag{2.3.35}$$

This allows one to combine the denominators of each of the propagator factors into a single fraction, which renders the otherwise quite inconvenient integrals that the Feynman calculus produces tractable. After gathering the denominators together in this fashion,

simplifying the result to remove redundant terms in $x$, and changing integration variable to isolate the $q$ contribution, one obtains

$$\mathrm{i}\hbar^{-1}\Sigma_2(k) = \frac{g_3^2}{36}\int_0^1 \mathrm{d}x \int \mathrm{d}^4 q \, \frac{1}{(q^2 + m^2 - x(1-x)k^2)^2}. \tag{2.3.36}$$

If this integral were being carried out over a Riemannian background, then one could exploit the spherical symmetry of the integrand to expand the volume measure as $2\pi^2 q^3 \, \mathrm{d}q$ after which the integral can be evaluated in terms of elementary functions. However, this is not the case here. The integral is to be carried out over spacetime, which carries an indefinite metric and prevents the simple expansion of the volume measure in this way. On the other hand, there is no real impediment to rotating the integration contour of $q_0$ (the time component of $q$) onto the imaginary axis, so that the integral is carried over over an honest Riemannian metric.

In order to properly justify this step, it is necessary to show that the contour does not cross any singularities of the integrand as it rotates. At first sight, it may appear that we step on a potential singularity at $q_0 = \pm\sqrt{\mathbf{q}^2 + m^2}$. However, the introduction of a small term $-\mathrm{i}\varepsilon$ into the denominator, where $\varepsilon$ is of square zero, shifts these poles to just below the positive real axis and just above the negative real axis. In rotating the contour, we do not pick up either of these poles, and the limit $\varepsilon \to 0$ can be taken at the end of the calculation. In fact one should introduce these $-\mathrm{i}\varepsilon$ terms already in (2.3.30) in order to define how to handle the integrand near the poles, but in the present treatment we have omitted such terms for clarity and worked at a purely formal level. In either case, the result is the same. After the rotating the contour, one is left with

$$\mathrm{i}\hbar^{-1}\Sigma_2(k) = \frac{\mathrm{i}g_3^2\pi^2}{18}\int_0^1 \mathrm{d}x \int_0^\infty \frac{q^3 \, \mathrm{d}q}{(q^2 + m^2 - x(1-x)k^2 - \mathrm{i}\varepsilon)^2}. \tag{2.3.37}$$

Since $q^2$ is now always positive, in virtue of the Riemannian metric, the $-\mathrm{i}\varepsilon$ terms can be dropped. This process is known as a Wick rotation. In carrying it out, we have acquired an overall factor of i from relating the Riemannian volume measure to the Lorentzian measure.

Unfortunately, as we stated earlier, this integral is manifestly divergent. This can be seen simply by counting powers in the integrand; since both the numerator and denominator in (2.3.37) behave like $q^4$, there is a logarithmic divergence at high $q$. This type of divergence, coming from large $|q|$ or small wavelengths, is known as an ultra-violet divergence, or just UV divergence for short. The alternative type, where the divergence comes from the $|q| \sim 0$ region of the integral, is called an infra-red divergence. By and

large, infra-red divergences are benign and can be handled without sophisticated surgery. Ultra-violet divergences are more malignant.[8] The ultra-violet divergence we have found here is similar to the divergence encountered when evaluating functional determinants in Section A.2.1, and although this divergence can also be sanitized using $\zeta$-function methods, it is more instructive to use an alternative method, known as dimensional regularization, due to 't Hooft and Veltman.

One possibility is simply to cut off the domain of integration in (2.3.37) at some large but finite momentum $\Lambda$. This would correspond to the introduction of a smallest length scale of order $1/\Lambda$ below which quantum fluctuations would not exist. In fact, this idea has a good deal of merit. Many quantum field theories are not exact theories of nature, but effective theories that describe the collective dynamics of some physical system within a limited range of validity. In many cases, this range of validity is such that the theory provides a good approximation below some energy or momentum scale $\Lambda$, above which new physics enters the problem. For example, new physics associated with quantum gravity is expected to appear at the Planck scale $M_{\rm P}$, in which case one should not trust field theory to provide a good approximation at momenta $q$ larger than the Planck scale, and (2.3.37) should be cut off at $\Lambda \sim M_{\rm P}$. This is the only choice for completely general quantum field theories, but most theories we deal with at low energy are rather better behaved than this and do not depend on the details of interactions at high energies. For such well-behaved ("renormalizable") theories, the method of 't Hooft and Veltman is very satisfactory.

---

[8]Until very recently it was a common conjecture, although without proof, that string theory does not contain any ultra-violet divergences. This is one of the attractive features that makes string theory an interesting proposition to workers in the field. The difficulty in proving ultra-violet convergence stemmed from the fact that techniques for handling high orders in perturbation theory were not well developed. A few years ago, a proof was offered (D'Hoker and Phong, 2002a,b,c,d) that superstrings were finite to two loops, but the generalization of the techniques used in the proof to an arbitrary number of loops was not straightforward. The situation seems finally to have been resolved in a very recent paper (Berkovits, 2004) which supplies a proof that string theory is finite to all orders in perturbation theory. This proof is not a generalization of the results of D'Hoker and Phong, but rather based on entirely new physics, an apparent duality between weakly coupled gauge theory and string theory in twistor space (Witten, 2003). This duality, if true, completes the connexion between gauge theory and string theory by providing a gauge theory dual in the weakly coupled régime that the other dualities, to be discussed in Chapter 3, do not probe.

The degree of divergence present in (2.3.36) is dependent on the dimension of spacetime. In $d$ dimensions, in fact, the integral has the form

$$\begin{aligned} \mathrm{i}\hbar^{-1}\Sigma_2(k) &= \frac{\mathrm{i}g_3^2\pi^{d/2}}{18\Gamma(d/2)}\int_0^1 \mathrm{d}x \int_0^\infty \frac{q^{d-1}\,\mathrm{d}q}{(q^2+m^2-x(1-x)k^2)^2} \\ &= \frac{\mathrm{i}g_3^2\pi^{d/2}}{36}\Gamma(2-d/2)\int_0^1 \mathrm{d}x\ \left[m^2-x(1-x)k^2\right]^{d/2-2} \end{aligned} \tag{2.3.38}$$

where the $q$-integral can be evaluted, as shown, in terms of the $\Gamma$-function if $d \neq 4$. If one attempts to analytically continue this expression to $d = 4$ then the pole of the $\Gamma$-function at $\Gamma(0)$ reproduces the divergence present in the unregulated integral (2.3.37). In detail, setting $\varepsilon = 2 - d/2$, then the self-energy satisfies

$$\mathrm{i}\hbar^{-1}\Sigma_2(k) = \frac{\mathrm{i}g_3^2}{72}\frac{\pi^2}{\varepsilon}\lim_{\varepsilon\to 0}\int_0^1 \mathrm{d}x\ \left[m^2-x(1-x)k^2\right]^{-\varepsilon}. \tag{2.3.39}$$

Expanding the integrand in a Taylor series and fixing the renormalized mass $m_R$ by asking that the dressed propagator has a pole at $k^2 = -m_R^2$ gives the $O(g_3^2)$ renormalized mass as

$$m_R^2 = m^2 + \frac{g_3^2\pi^2\hbar}{72\varepsilon} - \frac{g_3^2\pi^2\hbar}{72}\int_0^1 \mathrm{d}x\ \ln\left[m_R^2(1+x-x^2)\right], \tag{2.3.40}$$

where in the radiative term we have replaced $m$ by $m_R$, which is correct to the order to which we are working. To give a finite answer, we choose the bare mass to cancel the pole term, which is proportional to $1/\varepsilon$, together with a finite piece $m_b$ which is to be determined by[9]

$$m_b^2 = m_R^2 + \frac{g_3^2\pi^2\hbar}{72}\left(\frac{\pi}{\sqrt{3}} - 2 + 2\ln m_R\right). \tag{2.3.41}$$

**2.3.4. The quantum effective potential.** In (2.3.20), we wrote the partition function $Z[J]$ for a theory with action $S[\phi]$, and noticed that it was the vacuum–vacuum amplitude for the theory in the presence of an external current $J$. The partition function can be made the basis of a rather elegant treatment of radiative corrections, which we shall apply in Chapter 6.

The partition function $Z[J]$ includes all vacuum–vacuum diagrams, including connected and disconnected graphs. In general, a diagram which contains $N$ components will contribute to $Z[J]$ a term which involves the product of the $N$ connected graphs, with a symmetry factor $1/N!$ to take account of permutations of the diagram which simply exchange one connected component for another. Bearing this in mind, there is a convenient

[9]Notice that the radiative term is a first order correction in the deformation parameter $\hbar$. This is quite generally true. The loop expansion is an expansion in powers of $\hbar$.

factorization of $Z[J]$, in terms of the sum $\mathrm{i}\hbar^{-1}W[J]$ of all connected vacuum–vacuum graphs,

$$Z[J] = \sum_{N=0}^{\infty} \frac{1}{N!} \left( \frac{\mathrm{i}}{\hbar} W[J] \right)^N = \exp\left( \frac{\mathrm{i}}{\hbar} W[J] \right). \tag{2.3.42}$$

The quantum effective action is defined as a Legendre transform of $W[J]$,

$$\Gamma[\phi] = W[\tilde{J}] - \int \mathrm{d}v\, \phi(x)\tilde{J}(x), \tag{2.3.43}$$

where $\mathrm{d}v$ is the appropriate volume measure ($\mathrm{d}^4x$ in four-dimensional flat space) and $\tilde{J}$ is the current required to produce a mean field $\langle \phi(x) \rangle_{J=\tilde{J}} = \phi(x)$,

$$\begin{aligned} \phi(x) = \langle \phi \rangle_{J=\tilde{J}} &= -\frac{\mathrm{i}\hbar}{Z[\tilde{J}]} \frac{\delta}{\delta J(x)} Z[J] \Big|_{J=\tilde{J}} \\ &= \frac{\delta W[J]}{\delta J(x)} \Big|_{J=\tilde{J}}. \end{aligned} \tag{2.3.44}$$

The reason for the term quantum effective action is that $\Gamma[\phi]$ *is* the action for the field $\phi$, with quantum corrections such as the radiative terms discussed in the previous section already taken into account (Rivers, 1988; Weinberg, 1994). To see this in detail, one need only differentiate $\Gamma[\phi]$ with respect to $\phi$, which gives

$$\frac{\delta \Gamma[\phi]}{\delta \phi(x)} = \int \mathrm{d}v\ \frac{\delta W[J]}{\delta J(x)} \Big|_{J=\tilde{J}} \frac{\delta \tilde{J}[\phi(x)]}{\delta \phi(y)} - \int \mathrm{d}v\ \phi(x) \frac{\delta \tilde{J}[\phi(x)]}{\delta \phi(y)} - \tilde{J}[\phi]. \tag{2.3.45}$$

However, using the definition of $\tilde{J}$, the integral terms just cancel out, leaving only

$$\frac{\delta \Gamma[\phi]}{\delta \phi(x)} = -\tilde{J}. \tag{2.3.46}$$

In order to calculate values in the original quantum theory of interest, one sets the currents $J$ to zero, $J = 0$, which immediately shows that $\delta\Gamma/\delta\phi = 0$. Moreover, the sum $\mathrm{i}W[J]$ of connected graphs can be calculated as a sum of tree graphs with vertices and propagators taken from the action $\Gamma[\phi]$ rather than the bare action $S[\phi]$ (Weinberg, 1994). All these considerations dictate that $\Gamma[\phi]$ really is the action for the theory, with quantum effects built in.

Taking all this into account, the coefficients of any functional Taylor expansion of $\Gamma[\phi]$ in terms of the fields around some fixed configuration $\phi = \phi_0$ should be regarded as the *renormalized* coupling constants, with the renormalization energy scale set by the configuration point $\phi_0$ rather than by a mass-shell condition, as was done in the previous section. Since it includes all quantum corrections, one might imagine that $\Gamma[\phi]$ is a difficult

quantity to compute directly, but this is not the case. Consider the partition function $Z_\eta[J]$ calculated with a shifted action,

$$Z_\eta[J] = \int [\mathrm{d}\phi] \exp\left(-\frac{\mathrm{i}}{\hbar} S[\phi+\eta] + \frac{\mathrm{i}}{\hbar}\int \mathrm{d}v\; J\phi\right). \tag{2.3.47}$$

By changing variable in the path integral, this is just the same as

$$Z_\eta[J] = Z[J] \exp\left(-\frac{\mathrm{i}}{\hbar}\int \mathrm{d}v\; J\eta\right). \tag{2.3.48}$$

Since $\mathrm{i}\hbar^{-1}W[J] = \ln Z[J]$, the sum of all connected vacuum–vacuum diagrams calculated with $Z_\eta[J]$ instead of $Z[J]$ satisfies

$$W_\eta[J] = W[J] - \int \mathrm{d}v\; J\eta. \tag{2.3.49}$$

Recalling the definition of the quantum effective action (2.3.43), it is clear that $\Gamma[\phi]$ can be computed by no more than the sum of all connected diagrams, for which the diagram cannot be disconnected by cutting through one internal line, but calculated with a shifted action:

$$\Gamma[\phi_0] = \mathrm{i}\hbar \int_{\substack{\text{1PI}\\ \text{connected}}} [\mathrm{d}\phi] \exp\left(-\frac{\mathrm{i}}{\hbar} S[\phi+\phi_0]\right). \tag{2.3.50}$$

Such diagrams are called one-particle irreducible diagrams, or 1PI.

In cases of high symmetry, there is a further simplification which can be made in the formalism. Suppose we wish to calculate $\Gamma[\phi_0]$ for a position-independent field, where $\phi_0(x) = \phi_0$, say. Each term in the quantum effective action will contain a factor of the volume of spacetime, which arises from integrating over the position-independent field. Once this has been factored out, we write

$$\Gamma[\phi_0] = -\operatorname{Vol}(M_4) V^*(\phi_0) \tag{2.3.51}$$

where Minkowski space $M_4$ should be replaced by the appropriate volume of the background manifold, and $V^*$ is called the quantum effective potential.

## 2.4. Quantization of gauge field theories

The final piece of standard technology we shall need to apply is the quantization of Yang–Mills fields. These are field theories in which the Lagrangian is invariant under the action of some semi-simple Lie group $G$. As strictly understood, the term Yang–Mills refers only to the case where $G$ is non-Abelian, but the Abelian case is also physically interesting and in fact provides a description of low-energy electromagnetism. The term gauge theory

covers both Abelian and non-Abelian cases. Important examples are an $SU(2) \times U(1)$ theory, which describes the electroweak force; $SU(3)$, which describes quantum chromodynamics; and, as we shall see, $GL(d)$, which describes $d$-dimensional gravity.

Consider some action $S = \int L[q, \partial q]$ defined by a Lagrangian form $L$, where $q$ is regarded as a generalized coordinate which may be, for example, a quantum field. In the examples we have been dealing with up to now, $L$ has the form $L = \frac{1}{2}\phi\triangle\phi$ plus interaction terms, or if more than one field is present then $L$ should have the form

$$L = \frac{1}{2}\phi^i \triangle_{ij} \phi^j + \text{interaction terms}, \tag{2.4.1}$$

where $\triangle_{ij}$ is usually diagonal. Carrying out path integrals in this theory always involves the propagators $\triangle_{ii}^{-1}$ (no sum) for each field $i$, but in order for these to exist, $\triangle$ must be invertible. This implies that $\triangle$ is non-degenerate: $\triangle\phi = 0$ only if $\phi = 0$. However, in a gauge theory $L$ is supposed to be invariant under the action of some Lie group $G$. Therefore acting with $\triangle$ on $G$-related fields $\phi$ and $\phi^g$ (say) produces the same result.[10] It follows that $\triangle$ is a projection operator on the orbits of $G$, and cannot have a unique inverse. To formalise this, one can consider the Hessian of the Lagrangian,

$$h_{ij} = \left.\frac{\partial^2 L}{\partial\phi^i \partial\phi^j}\right|_{\phi^k = 0}. \tag{2.4.2}$$

This is just equal to $\triangle_{ij}$ in the example above. If $\det h_{ij} = 0$, then the Lagrangian is said to be singular. The condition that the Lagrangian is non-singular is the same as the condition that the symplectic two-form $\omega$ is invertible (Fadeev, 1999), so until now we have implicitly been assuming that the Lagrangian is non-singular. In theories of a single scalar field this is almost always the case, whereas for gauge theories it is not, the offending singularity giving rise to difficulties in quantization. The solution of these difficulties, found in the late 1960s by Yang & Mills (with contributions by Feynman and de Witt) involves the study of mechanics with constraints.

[10] We are temporarily ignoring the possibility that acting with an element of $G$ on $L$ gives back not just $L$ but also a total derivative which decays sufficiently fast at the boundary or at infinity that the total action is unchanged. For example, this is always the case when the theory is invariant under supersymmetry and also for Galilei transformations in classical mechanics. Such Lagrangians are called quasi-invariant. If a Lagrangian is quasi-invariant, then there is no way to remove the offending term just by adding a total derivative to $L$ (which classically would define the same theory). The underlying cause is always non-trivial cohomology of the group $G$ (de Azcárraga and Izquierdo, 1995).

**2.4.1. Constrained Hamiltonian systems and reduced phase space.** Consider some phase space $M$ with local coordinates $\xi^m$, not necessarily symplectic, and a general Lagrangian form (on $M \times \mathbf{R}$)

$$L = f_m(\xi)\,\mathrm{d}\xi^m - \phi(\xi)\,\mathrm{d}t = \gamma - \phi\,\mathrm{d}t. \tag{2.4.3}$$

The Lagrangian may be singular or non-singular; this is a sufficiently general framework to support either. The symplectic form induced by the Lagrangian is $\omega = \mathrm{d}\gamma$. Since $L$ can be singular, there is no requirement that $\omega$ be non-degenerate. Hamilton's equations in this form are

$$\omega_{mn}\dot{\xi^n} = \frac{\partial \phi}{\partial \xi^n}. \tag{2.4.4}$$

The Darboux theorem guarantees the existence of coordinates $(p_i, q^j, \lambda^\alpha)$ on $M$ such that $\omega$ can be written

$$\omega = p_i\,\mathrm{d}q^i + \mathrm{d}\theta(\lambda), \tag{2.4.5}$$

where $\theta$ is a local function of the $\lambda^\alpha$ which we assume it is permissible to discard when appearing under an integral, to give the action. We can expand the function $\phi(p, q, z)$ as a power series in the $\lambda^\alpha$,

$$\phi(p, q, z) = \phi(p, q) + \lambda^\alpha \varphi_\alpha(p, q) + \cdots. \tag{2.4.6}$$

Of course, if the Lagrangian is non-singular then the total derivative term $\mathrm{d}\theta$ is not present and the Taylor series contains only the term $\phi(p, q)$, which is the Hamiltonian. Combining all of these elements, the Lagrangian form can quite generally be written

$$L = p_i q^i - h(p, q)\mathrm{d}t - \lambda^\alpha \varphi_\alpha(p, q). \tag{2.4.7}$$

In expressing the Lagrangian in this way we are assuming any $\lambda^\alpha$ which appear in terms of order quadratic or above are to be discarded. This is perfectly legitimate, because for such coordinates the field equation is an algebraic equation in $\lambda^\alpha$ (by virtue of the absence of derivatives of the $\lambda^\alpha$); we assume that such field equations have been solved, and the results inserted into $L$. Only if some $\lambda^\alpha$ enters at most linearly does its field equation not involve $\lambda^\alpha$; in such cases, the $\lambda^\alpha$ are not determined and play the role of Lagrange multipliers. The associated coefficient functions $\varphi_\alpha$ are called constraints since the variational equations

for the $\lambda^\alpha$ enforce $\varphi_\alpha(p,q) = 0$. Motion on phase space is not free but must satisfy the constraints. The reduced phase space is defined to be the manifold $\Gamma$ satisfying

$$\Gamma = \{(p,q) \mid \varphi_\alpha(p,q) = 0\}. \tag{2.4.8}$$

The Lagrange multipliers have been eliminated and the constraints are automatically satisfied on $\Gamma$. The reduced phase space may have a non-degenerate symplectic form, in which case it is possible to proceed immediately to quantization using the methods outlined above. If $\Gamma$ does not carry a non-degenerate symplectic form then the process can be repeated.

Two classes of constraints can be distinguished. A constraint is called first class if its Poisson bracket with all other constraints vanishes on $\Gamma$. Thus, if $\chi$ is a first class constraint then

$$[\chi, \varphi_\alpha]_{\mathrm{P}} = 0 \quad \text{for all } \varphi_\alpha \text{ on } \Gamma. \tag{2.4.9}$$

Constraints for which this is not true are called second class. If one writes the set of second class constraints as $\{\phi_\alpha\}$, then the matrix $C_{\alpha\beta}$ is defined by

$$[\phi_\alpha, \phi_\beta]_{\mathrm{P}} = C_{\alpha\beta} \tag{2.4.10}$$

The determinant $\det C_{\alpha\beta}$ is non-zero, so there must be an even number of second class constraints because an antisymmetric matrix of odd dimensionality always has a vanishing determinant. Since $C_{\alpha\beta}$ is non-singular, its inverse exists (in a matrix sense) and is denoted $C^{\alpha\beta}$. The goal is now to construct the reduced phase space $\Gamma$, but one must treat these general classes of constraints differently in the process. For each first class constraint $\chi$, one can solve for one of the coordinates or momenta in terms of the remaining $q^i$, $p_i$. This defines a submanifold $\tilde{M}$ of $M$ on which the constraint is satisfied, and is equivalent to setting the constraint to zero, $\chi = 0$. Because $\chi$ has zero Poisson bracket with the other constraints, this is consistent provided one also sets the canonical conjugate of $\chi$ to zero; then, $\chi$ has zero Poisson bracket with any local function on $\tilde{M}$. At this stage it is possible to discard $M$ and work on $\tilde{M}$, solving the next constraint which defines a further submanifold, and so on. If all constraints are first class, the submanifold one reaches at the end of this process is the reduced phase space. Eliminating each constraint reduces

the dimension of phase space by two, because one sets to zero both $\chi$ and its canonical conjugate.[11]

If second class constraints are present, on the other hand, then after the first class constraints have been exhausted one cannot continue this process of finding successive submanifolds. This happens because the second class constraints have non-zero Poisson brackets among themselves, so simply setting each second class constraint to zero is inconsistent on quantization. In order to take account of this, one defines a Dirac bracket

$$[f, g]^* = [f, g]_{\mathrm{P}} - [f, \varphi_\alpha]_{\mathrm{P}} C^{\alpha\beta} [\varphi_\beta, g]_{\mathrm{P}}. \tag{2.4.11}$$

This is sometimes called a starred bracket. It is easy to verify that the Dirac bracket of an observable with any constraint is zero, $[f, \varphi_\alpha]^* = 0$. Also, if either argument is first class then the Dirac bracket coincides with the Poisson bracket. If one now uses Dirac brackets to calculate commutators in the quantum theory, then it is consistent to set $\phi = 0$ by solving the constraint in the usual manner, but it is no longer necessary to set the canonical conjugate to zero also. Proceeding in this way yields the reduced phase space just as if no second class constraints had been present, except that each step reduces the dimension of phase space by one rather than two.

**Example.** As an illustration of this technique, consider the case of Maxwell theory in four dimensions (Fadeev, 1999). The theory consists of a $U(1)$ connexion $A = A_a \, \mathrm{d}x^a$ with curvature $F = \mathrm{d}A$ and the Lagrangian is $\mathcal{L} = F \wedge *F$. Schwinger wrote this Lagrangian as (in components)

$$\mathcal{L} = (\partial_a A_b - \partial_b A_a) F^{ab} - \frac{1}{2} F_{ab} F^{ab} \tag{2.4.12}$$

where $A_a$ and $F_{ab}$ are temporarily taken to be independent. One now splits $A$ and $F$ into timelike and spacelike pieces,

$$\mathcal{L} = (\partial_t A_j) F^{tj} + A_t \partial_j F^{0j} - F^{ij} (\partial_i A_j - \partial_j A_i) - \frac{1}{2} F_{0j} F^{0j} + \frac{1}{2} F_{ij} F^{ij} \tag{2.4.13}$$

(missing components from this Lagrangian, such as $F_{00}$, are zero in virtue of the antisymmetry of $F$). We identify $E^j = F^{0j}$ as the electric field and $\mathbf{A}$ as the vector potential. The quantity $F^{ij}$ is (using the Hodge star) the magnetic induction $\mathbf{H}$, which satisfies

$$\mathbf{H} = \operatorname{curl} \mathbf{A}. \tag{2.4.14}$$

[11] This whole discussion pre-supposes that the constraint submanifold $\tilde{M}$ has trivial topology. There is no reason why one cannot re-frame this procedure in terms of constraint hypersurfaces which have interesting topologies, but the result isn't relevant for our purposes.

One can now recognize the various components which appeared in the general discussion above. The magnetic induction $F_{ij}$ and the timelike component of the connexion $A_t$ appear without time derivatives. Of these, $A_t$ enters linearly, so it defines a constraint. The field equation for the magnetic induction can be solved, which just returns $F_{ij} = \partial_i A_j - \partial_j A_i$. Substituting into the Lagrangian gives

$$\mathcal{L} = (\partial_t A_j)E^j - \frac{1}{2}(\mathbf{E}^2 - \mathbf{H}^2) + A_0 \partial_j E^j. \tag{2.4.15}$$

The constraint defined by $A_0$ is Gauss' Law, $\varphi = \operatorname{div} \mathbf{E} = 0$. This constraint is first class, ie. $[\varphi(x), \varphi(y)]_{\mathrm{P}} = 0$. For any function $\lambda(y)$ on $\mathbf{R}^3$, one has

$$[A_j(x), \int \varphi(y)\lambda(y)\, \mathrm{d}^3 y]_{\mathrm{P}} = \partial_j \lambda(x). \tag{2.4.16}$$

This means that $\varphi$ generates gauge transformations of $\mathbf{A}$. The gauge group $\mathcal{G}$ is the exponentiation of $\varphi$, $\mathcal{G} = \{\mathrm{e}^{\mathrm{i}\theta\varphi} \mid \theta \in (0, 2\pi]\}$, which gives a different $U(1)$ at each point in spacetime. An element $\lambda \in \mathrm{Lie}(\mathcal{G})$ sends $\mathbf{A}$ to $\mathbf{A} + \partial\lambda$,

$$\delta\mathbf{A} = [\mathbf{A}, \lambda]_{\mathrm{P}} = \partial\lambda. \tag{2.4.17}$$

This shows the necessity of setting all Poisson brackets with $\varphi$ to zero. (Recall that this is only necessary with first class constraints.) This is equivalent to choosing a gauge for $\mathbf{A}$, and is a general rule in gauge theory: the constraints are generators of gauge transformations under the Poisson bracket. Setting this Poisson bracket to zero is equivalent to fixing a gauge for $\mathbf{A}$; a common choice is Coulomb gauge, in which $\operatorname{div} \mathbf{A} = 0$. Therefore the three degrees of freedom present in $\mathbf{A}$ are reduced to two: in physical terms, light has two polarizations, not three.

**2.4.2. The Fadeev–Popov determinant.** Following this procedure, in most cases, eventually leads to a non-singular Lagrangian that one can work with. The result is a theory in which the gauge invariance has been "de-parametrized" or lost: in other words, the gauge invariance under $G$ that was a defining property of the original theory is no longer manifest in the reduced phase space description. For many physical reasons (like the Ward identity), it is often convenient to try and retain the appearance of gauge invariance in our description. The Fadeev–Popov procedure is a way of doing this, although the result still depends on a set of gauge non-invariant functionals $\Lambda^\beta$ called the gauge-fixing functionals. Later we will introduce the BRST formalism, in which the independence of this procedure from $\Lambda^\beta$ is clear.

Let $\mathcal{L}$ by a Lagrangian invariant under some symmetry group $G$. The partition function of this theory should be defined by

$$Z = \frac{1}{\mathrm{Vol}(G)} \int [\mathrm{d}p][\mathrm{d}q] \, \exp\left(-\mathrm{i} \int_M \mathrm{d}v \, \mathcal{L}\right). \tag{2.4.18}$$

where $\mathrm{Vol}(G)$ is the volume of the gauge group, and $p$ and $q$ are coordinates on phase space. If desired, one can rotate to Euclidean signature. Eq. (2.4.18) just says that it is integrating along gauge orbits which causes difficulties in a perturbative evaluation of the functional integral. Dividing out the number of configurations which are related by gauge transformations, $\mathrm{Vol}(G)$, will remove this difficulty. The Fadeev–Popov procedure is a technique to actually evaluate (2.4.18) by replacing the integrand with an expression which evaluates to unity (rather than $\infty$) along orbits of $G$. Eq. (2.4.18) is then just equal to the integral of this expression. In particular, we could replace the entire integrand with a $\delta$-function of some gauge condition.

To see how this works in detail, we return to the description of a singular Lagrangian on a phase space $\Gamma$ with all excludable variables removed (that is, variables entering quadratically or above, but without time derivatives), as in (2.4.7). The physical phase space is the manifold[12]

$$\Gamma^* = \{(p,q) \in \Gamma \mid \varphi_\alpha(p,q) = 0\}/\mathcal{G}, \tag{2.4.19}$$

where $\mathcal{G}$ is the gauge group generated by the constraints $\varphi_\alpha$ through the Poisson bracket. In order to work with $\Gamma^*$ explicitly, it is necessary to introduce coordinates, so for this purpose one chooses an auxiliary gauge condition $\Lambda^\beta = 0$. This defines a hypersurface which should intersect each orbit of $\mathcal{G}$ exactly once: choosing such a hypersurface may require some care. In addition, the $\Lambda^\beta$ should commute under the Poisson bracket, so they behave like supplementary first class constraints.

In terms of the functional integral, we now wish to restrict attention to the submanifold of phase space defined by $\varphi_\alpha = 0$, $\Lambda^\beta = 0$. We take $p$ and $q$ to be decomposed as

$$q = (q^\alpha, q^*) \quad \text{and} \quad p = (p_\beta, p^*) \tag{2.4.20}$$

[12] $\tilde{\Gamma}$ is generically a manifold, but may have unusual properties. In particular, quotient spaces are often not Hausdorff.

where $(q^*, p^*)$ coordinatize the reduced phase space $\Gamma^*$, and $q^\alpha$, $p^\beta$ are supposed to solve $\Lambda^\alpha = 0$, $\varphi_\beta = 0$,

$$\begin{aligned} \Lambda^\alpha = 0 \quad &\text{implies} \quad q^\alpha = q^\alpha(q^*, p^*) \\ \varphi_\beta = 0 \quad &\text{implies} \quad p_\beta = p_\beta(q^*, p^*). \end{aligned} \tag{2.4.21}$$

One can write the Liouville measure on $\Gamma^*$ using

$$\begin{aligned} [\mathrm{d}p][\mathrm{d}q] &= [\mathrm{d}p_\beta][\mathrm{d}p^*][\mathrm{d}q^\alpha][\mathrm{d}q^*] \\ &= \frac{\partial(p_\beta, q^\alpha)}{\partial(\varphi_\gamma, \Lambda^\delta)} [\mathrm{d}\Lambda^\delta][\mathrm{d}p^*][\mathrm{d}\varphi_\gamma][\mathrm{d}q^*]. \end{aligned} \tag{2.4.22}$$

Moving the Jacobian to the left hand side and restricting the support of the measure to the constraint and gauge-fixing hypersurfaces shows that

$$[\mathrm{d}p^*][\mathrm{d}q^*] = \delta(\Lambda)\delta(\varphi) \, |\det[\varphi, \Lambda]_{\mathrm{P}}| \, [\mathrm{d}p][\mathrm{d}q]. \tag{2.4.23}$$

After this change of variables, the partition function becomes

$$Z = \int [\mathrm{d}p][\mathrm{d}q][\mathrm{d}\lambda] \, \delta(\Lambda) \, |\det[\varphi, \Lambda]_{\mathrm{P}}| \, \exp\left[-\mathrm{i}\int_M \mathrm{d}v \, (p\dot{q} - H(p,q) - \lambda^\alpha \varphi_\alpha)\right], \tag{2.4.24}$$

where we have used the identity $\delta(\varphi) = \int [\mathrm{d}\lambda] \, \mathrm{e}^{\mathrm{i}\lambda\phi}$ (up to a constant) to make use of the constraint terms already present in the Lagrangian. To summarise, this says that one can evaluate the path integral for the non-singular gauge-fixed Lagrangian on the reduced phase space $\Gamma^*$, with all constraints applied, by working with the singular, gauge-invariant Lagrangian and the full phase space, provided one includes a gauge-fixing $\delta$-function and the Fadeev–Popov determinant $\partial(\varphi, \Lambda)/\partial(p, q)$. This last term is sometimes written $\Delta_{\mathrm{FP}}$.

The primary obstacle to a perturbative evaluation of (2.4.18), namely the singularity of the Lagrangian, has thus been removed. Two difficulties still remain: the $\delta$-function itself, and $\Delta_{\mathrm{FP}}$. To produce a more manageable expression for the $\delta$-function, let $\Lambda_\alpha$ be given by

$$\Lambda_\alpha = f_\alpha - \omega_\alpha \tag{2.4.25}$$

where $f_\alpha$ is a local function of the fields and $\omega_\alpha$ is any prescribed function, not depending on the fields. For example, in electrodynamics a common choice for $f$ is the Lorentz condition $\partial_a A^a$. This is the choice we shall actually use later, in Section 6.4. Setting $\Lambda$ as above then gives a kind of generalized Lorentz gauge, but the partition function itself (2.4.18)

does not depend on $\omega$. If one then integrates over $\omega$ with a Gaussian smearing function, one obtains

$$Z \int [\mathrm{d}\omega] \exp\left(-\frac{\mathrm{i}}{2\xi}\int_M \mathrm{d}v\, \omega_\alpha \omega^\alpha\right) = \int [\mathrm{d}p][\mathrm{d}q][\mathrm{d}\lambda]\, |\det[\varphi,\Lambda]_{\mathrm{P}}|\, \exp\left[-\mathrm{i}\int_M \mathrm{d}v\, \left(\mathcal{L} + \frac{1}{2\xi} f_\alpha f^\alpha\right)\right] \tag{2.4.26}$$

where we have used the $\delta$ function to perform the $\omega$ integral on the right-hand side, and the integral $\int [\mathrm{d}\omega] \mathrm{e}^{-\mathrm{i}\omega^2/2\xi}$ on the left-hand side is an irrelevant overall constant that can be discarded. Therefore, the presence of the $\delta$-function is equivalent to adding an extra term $f_\alpha f^\alpha/2\xi$ to the action, where $\xi$ is an arbitrary parameter. Such an action is usually said to be in a generalized $\xi$-gauge.

$\Delta_{\mathrm{FP}}$ depends on the choice of $\Lambda$. Recall that the constraints $\varphi_\alpha$ are generators of infinitesimal gauge transformations, ie.,

$$\delta\Lambda^\alpha = [\Lambda^\alpha, \int \lambda^\beta(y)\varphi_\beta(y)\, \mathrm{d}^3 y]_{\mathrm{P}}. \tag{2.4.27}$$

Bearing this in mind, the determinant in $\Delta_{\mathrm{FP}}$ is the same as

$$|\det[\varphi,\Lambda]_{\mathrm{P}}| = \left|\det \frac{\partial(\Lambda^\alpha)^{g(\lambda)}}{\partial\lambda^\beta}\right|, \tag{2.4.28}$$

where $\Lambda^{g(\lambda)}$ is the result of applying a small gauge transformation with parameters $\lambda$ to $\Lambda$. This provides a useful way of calculating the determinant. For example, in the case where the gauge-fixing function is the Lorentz-like combination $\mathcal{D}_a A^a$, one obtains

$$\Delta_{\mathrm{FP}} = \det \Box_{\mathrm{FP}} = \det\left(\partial^a\partial_a + \partial^a[A_a, \cdot]\right) = \det \mathcal{D}_a\mathcal{D}^a. \tag{2.4.29}$$

(Of course, this is only up to $\delta$-functions which correspond to the identity operator in coordinate space, and which we have discarded.) Evidently, in the present case $\Box_{\mathrm{FP}}$ is the gauge-covariant derivative derived earlier, in Section B.5.2. Once expressed in this form, the determinant can be rewritten as a Gaussian integral, using the Grassmann (or Berezin) identity

$$\det A \propto \int [\mathrm{d}c][\mathrm{d}\bar{c}]\, \exp\left(-\frac{\mathrm{i}}{2}\int_M \mathrm{d}v\, \bar{c}\Box_{\mathrm{FP}} c\right). \tag{2.4.30}$$

The fields $c$ and $\bar{c}$ which appear here are taken to be anticommuting, and since they are scalars they violate the spin-statistics theorem. Nonetheless, there is nothing really amiss here: the spin-statistics theorem is a statement about which types of fields may appear in 'in' and 'out' scattering states defined at asymptotically early and late time, and it is easy to see (for example by deriving the Feynman rules) that states containing $c$ and $\bar{c}$

particles do not appear as in or out states. The particles created and annihilated by $c$ and $\bar{c}$ are therefore called ghosts; they are absent at tree level, and circulate only in loops. Substituting this into the Lagrangian shows that the final path integral which one should evaluate is

$$Z = \int [\mathrm{d}p][\mathrm{d}q][\mathrm{d}z][\mathrm{d}c][\mathrm{d}\bar{c}] \, \exp\left[-\mathrm{i}\int_M \mathrm{d}v \, \left(\mathcal{L} + \frac{1}{2\xi} f^\alpha f_\alpha + \frac{1}{2}\bar{c}\Box_{\mathrm{FP}} c\right)\right] \tag{2.4.31}$$

The Feynman rules, including rules for the ghosts, can be read off from this Lagrangian. In this state, the integral can be calculated perturbatively as was desired.

## 2.5. BRST symmetry

Although the Fadeev–Popov technology described above represents the most common approach to concrete calculations in gauge theories, questions of principle are usually more accessible using a somewhat refined technique known as BRST quantization.[13] In this formalism, the invariance of the quantization procedure from the gauge-fixing functional $\Lambda^\alpha$ employed in the previous section is manifest. Moreover, there is a well defined procedure based on cohomological methods (see Appendix B) for deciding which states in the quantum gauge theory are physical and which are not. Historically, BRST invariance was very important in proving the renormalizability of non-Abelian gauge theories. Although we are content with the somewhat more restrictive Fadeev–Popov framework for actual calculations (Chapter 6), it is worthwhile to briefly describe the BRST method here to complete the story of quantization of gauge-invariant quantum theories, and also to set the scene for the quantization of the relativistic string to be undertaken in the next chapter.

The starting point for BRST quantization is the gauge-fixed Lagrangian (essentially (2.4.31)) which has the form

$$\mathcal{L}_{\mathrm{FP}} = \mathcal{L}(\phi) + B(f) + \underbrace{b_\alpha c^A \delta_A \Lambda^\alpha}_{\text{FP term}} + \underbrace{f_\alpha \Lambda^\alpha}_{\delta_{\mathrm{D}}\text{ piece}}, \tag{2.5.1}$$

in which the original action $\int \mathcal{L}$, which is taken to be a function of a set of matter and gauge fields $\phi_i$, obeys a symmetry generated by $\delta_A$ with parameters $\varepsilon^A$,

$$\delta\mathcal{L} = -\varepsilon^A \delta_A \mathcal{L}. \tag{2.5.2}$$

[13] the initials BRST are associated with its discovers, Becchi, Rouet, and Stora (1975) and Tyutin (1975).

The generators $\delta_A$ are supposed to obey some Lie algebra $[\delta_A, \delta_B] = f^C_{AB}\delta_C$ and we have introduced yet another representation of the Fadeev–Popov determinant as the combination $b_\alpha c^A \delta_A \Lambda^\alpha$ in which $b_\alpha$ and $c^A$ constitute the ghost fields. The term $f_\alpha \Lambda^\alpha$ merely enforces the $\delta$-functional $\delta_{\rm D}(\Lambda^\alpha)$, and, if desired, one can include a non-trivial function $B(f)$ of the $f_\alpha$ corresponding to a Gaussian-averaged (generalized-$\xi$) gauge, or indeed any other gauge-fixing term one likes. There is no real loss of generality in taking $B(f) = 1$, which corresponds to dealing with the $\delta$-functional directly and not using an averaged gauge.

Define the BRST transformation by

$$\begin{aligned} \delta_\theta \phi_i &= -\theta c^A \delta_A \phi_i \\ \delta_\theta b_\alpha &= \theta f_\alpha \\ \delta_\theta f_\alpha &= 0 \\ \delta_\theta c^A &= -\frac{\theta}{2} f^A_{BC} c^B c^C, \end{aligned} \tag{2.5.3}$$

where $\theta$ is an infinitesimal parameter which must be taken to be anticommuting, since $\delta_\theta$ mixes even and odd variables. When acting on any gauge-invariant functional of the matter fields $\phi_i$, such as the original Lagrangian, the BRST variation $\delta_\theta$ is zero, since $\delta_\theta \phi_i$ is just a gauge transformation with infinitesimal parameter $-\theta c^A$.[14] Clearly $\delta_\theta$ annihilates $B(f)$, and when acting on the remaining terms one obtains

$$\begin{aligned} \delta_\theta \left[ b_\alpha c^A \delta_A \Lambda^\alpha + f_\alpha \Lambda^\alpha \right] = \; & \theta f_\alpha c^A \delta_A \Lambda^\alpha + b_\alpha \frac{\theta}{2} f^A_{BC} c^B c^C \delta_A \Lambda^\alpha \\ & - b_\alpha c^A \theta c^B \delta_B \delta_A \Lambda^\alpha - f_\alpha \theta c^A \delta_A \Lambda^\alpha. \end{aligned} \tag{2.5.4}$$

The first and last terms cancel, whereas using the Lie algebra of $\delta_A$, one can evaluate the term containing two factors of $\delta_A$,

$$c^A c^B \delta_A \delta_B - c^A c^B \delta_B \delta_A = c^A c^B f^C_{AB} \delta_C \quad \text{so} \quad c^A c^B \delta_B \delta_A = -\frac{1}{2} c^A c^B f^C_{AB} \delta_C; \tag{2.5.5}$$

on substitution in (2.5.4) one can then show that the total variation of the gauge-fixing terms is zero. In consequence, $\delta_\theta \mathcal{L}_{FP}$ is zero and the gauge-fixed Lagrangian is BRST-invariant.

[14] This may, in fact, only hold up to boundary terms which vanish on integration. Such Lagrangians are called quasi-invariant and do occur in interesting physical theories. For example, there is no such thing as a supersymmetry-invariant Lagrangian: only a supersymmetry quasi-invariant Lagrangian.

Moreover, one can show by repeated application of the BRST transformations (2.5.3) applied to various combinations of the elementary fields that the BRST transformation is nilpotent (Weinberg, 1994), in the sense that

$$\delta_{\theta'}\delta_\theta F(\phi_i, b_\alpha, f_\alpha, c^A) = 0, \tag{2.5.6}$$

where $F(\phi_i, b_\alpha, f_\alpha, c^A)$ is any function of the elementary fields (including matter and gauge fields and the ghosts). The concept of the cohomology of a nilpotent operator has been introuduced in Appendix B (see also de Azcárraga and Izquierdo (1995)). Since the gauge-fixed Lagrangian $\mathcal{L}_{FP}$ is BRST-invariant, it follows that it can be written in the form (Weinberg, 1994)

$$\mathcal{L}_{FP} = \mathcal{L} + \delta_\theta \Psi \tag{2.5.7}$$

where $\Psi$ is some function of the fields in the theory, and $\mathcal{L}$ can be chosen to equal the original gauge-invariant Lagrangian. This identification suggests that the physical content of a gauge-fixed theory such as (2.4.31) or (2.5.1) can be identified with the cohomology of $\delta_\theta$, that is, a kernel term such as $\mathcal{L}$ which is not BRST-exact, in the sense that it can be written as a BRST transformation applied to some local function of the fields, and a BRST-exact term which contains details of the gauge-fixing contributions. The physical content is not supposed to depend on the gauge-fixing parts, containing terms such as $\Lambda^\alpha$ which appeared to be arbitrary choices in the Fadeev–Popov formalism. One might now hope to show that theories lying in the same BRST cohomology class give rise to the same physical predictions, which frees the formalism from any dependence on $\Lambda^\alpha$.

To make this idea precise, consider any matrix element $\langle\alpha|\beta\rangle$ between two states $|\alpha\rangle$ and $|\beta\rangle$. If the scattering amplitude or S-matrix element or other physical observable defined by $\langle\alpha|\beta\rangle$ is to be invariant under changes in the choice of gauge-fixing functional – in terms of the BRST representation, this is more accurately states as invariance under choices of the BRST-exact term $\delta_\theta\Psi$ – then the variation $\delta\langle\alpha|\beta\rangle$ under a change $\delta\Psi$ in $\Psi$ must vanish,

$$\delta\langle\alpha|\beta\rangle = \mathrm{i}\langle\alpha|\delta_\theta\delta\Psi|\beta\rangle = 0 \tag{2.5.8}$$

(working to first order in $\delta\Psi$). If one introduces a fermionic or odd BRST charge $Q$, defined as usual via the commutation or anticommutation rule

$$\delta_\theta\varphi = \mathrm{i}\theta[Q, \varphi]_\mp \tag{2.5.9}$$

where the quantity $[Q, \varphi]_{\mp}$ is a commutator or anti-commutator depending on whether $\varphi$ is even or odd ($Q$ itself being odd). Therefore,

$$\delta_\theta \langle \alpha | \beta \rangle = \theta \langle \alpha | [Q, \delta\Psi] | \beta \rangle = 0. \tag{2.5.10}$$

One can show that nilpotence of $\delta\theta$ implies that $Q$ is also nilpotent, or that $Q^2 = 0$. If (2.5.10) is to vanish for all choices of $\delta\Psi$, then the charge $Q$ applied to each of $|\alpha\rangle$ and $|\beta\rangle$ must vanish individually,

$$Q|\alpha\rangle = 0 \quad \text{and} \quad Q|\beta\rangle = 0. \tag{2.5.11}$$

This is fundamental requirement of physical states: they are in the kernel of the BRST charge operator $Q$. Therefore, any states of the form

$$|\alpha\rangle + Q|\Psi\rangle \tag{2.5.12}$$

are physically equivalent to $|\alpha\rangle$, because when taken in matrix elements or inner products associated with physical observables, terms of the form $Q|\Psi\rangle$ must vanish. Thus, independent physical states correspond to the cohomology classes of $Q$.

If desired, one can define an entire quantization scheme based on these results, although we shall not find such a power tool necessary and therefore do not give details. Given a gauge-invariant Lagrangian $\mathcal{L}$, one takes the physical theory associated with $\mathcal{L}$ to be the combination

$$\mathcal{L}_{\text{BRST}} = \mathcal{L}(\phi_i) + \delta_\theta \Psi(\phi_i, b_\alpha, f_\alpha, c^A) \tag{2.5.13}$$

where $\Psi$ is the most general function of the matter and ghost fields consistent with symmetries and local gauge invariances of the theory (Weinberg, 1994). One can show that the S-matrix states annihilated by $Q$ and the decoupling of ghosts from asymptotic 'in' and 'out' states do not depend on the choice of gauge-fixing term $\Psi$. (However, it does appear necessary to prove unitarity of the S-matrix, and that the ghosts do indeed decouple in any individual theory.)

CHAPTER 3

# String theory, compactification, and membranes

As a final explanatory topic before moving on to the cosmological theories which are the principal focus of this thesis, we devote some time to explaining the background in string theory which supports membrane models. It is not necessary to go very far into string theory to provide sufficient background. In fact, most of the calculations to be carried out in Part 2 do not involve a detailed understanding of string theory at all, and can be carried out only using standard techniques in general relativity and quantum field theory. Therefore, the technical constructions described in this chapter will not reappear later, and most of this chapter can be read independently of the others or be skipped if desired. Nonetheless, since the principal focus of this thesis is cosmologies built on brane physics, it seems profitable and appropriate to sketch an outline of the supporting theory.

With this goal in mind, the emphasis in this presentation is rather different to conventional treatments. These are typically centred around a careful quantization of the string and an examination of the string spectrum (Green et al., 1987; Polchinski, 1998). More modern developments, relating to D-branes, brane physics, gravity/gauge theory dualities, and much else can be found in Johnson (2003). Here we focus more narrowly on how conventional gravity and particle physics theories are supposed to arise from string theory, in particular dealing with the string sigma-model and compactification to four or five dimensions. We then mention M-theory and move on to a discussion of the string dualities which result in the appearance of D-branes.

String theory is not deeper than quantum mechanics. Everything described in this chapter relies on the general principles of quantum mechanics which were set out in the last chapter, so string theory does not replace quantum mechanics or provide explanations for those features which, over the years, various workers and commentators have considered 'unsatisfactory', 'unphysical' or 'odd'. Instead, string theory relies on quantum mechanics in an essential way. The explanations of familiar phenomena which string theory provides relate almost entirely to the emergence of general relativity as a dynamical theory of mass at

low energies. The other observed interactions can be almost entirely explained satisfactorily explained by appealing to the gauge principle, without recourse to any deeper or more detailed theory. It is true that string theory does provide a home for gauge theories, holding out the hope that at some point in the future we may be able to explain the appearance of the Standard Model gauge group $SU(3) \times SU(2) \times U(1)$ and three fermion families from dynamical properties of the underlying string theory. But the deepest insights which have come out of string theory over the last three decades relate almost entirely to gravitational phenomena, and (very recently) specific examples of how gravity can be understood in certain régimes of parameter space as just another manifestation of a gauge theory.

Once the dust had settled on the final theory of non-relativistic quantum mechanics (outlined in Chapter 2) introduced by Schrödinger, Heisenberg, Dirac and others in the 1920s, it was clear that although the problems of atomic transitions which had first pointed the way to the quantum principle had largely been resolved, there remained some outstanding difficulties. At one level, there were processes in the quantum theory, such as $\beta$-decay, which might only be the work of time to explain, or might point to new underlying problems, but, more seriously, there were aspects of the classical theory – such as gravity, of whatever flavour, or electromagnetism – which had not been given a satisfactory quantum treatment. Above all, this quantum mechanics was not relativistic, nor was it obvious how to self-consistently introduce interactions into the theory.

The problem of gravity can be seen at the level of the Schrödinger equation. In general, this has the form

$$-\frac{\hbar}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi = -\mathrm{i}\hbar\frac{\partial\psi}{\partial t}, \tag{3.0.14}$$

where $V(x)$ is the classical potential. If one inserts the Newtonian gravitational binding energy, $V = \Phi$, one must express $\Phi$ via some expectation value of the wavefunction. For example, classically $\Phi = -GMm/r$, describing interaction of a particle with mass $m$ with some large mass $M$. More exactly, $\Phi$ should solve the Poisson equation,

$$\triangle\Phi = 4\pi G m|\psi|^2 \tag{3.0.15}$$

where $m|\psi|^2$ is the smeared mass density function for the particle; classically, this approaches a $\delta$-function. However, inserting the solution of this Poisson equation into the Schrödinger equation results in a non-linear wave equation, which is a contradiction. Quantum mechanics is an exactly linear theory, and combinations like $|\psi|^2$ are not allowed in

the fundamental equations of the theory, even though they represent physically meaningful quantities. At one level one can see that this must be true because the principal element utilised in the construction of quantum mechanics in the last chapter was the Hilbert space of states. The wavefunction $\psi(x)$ is in one-to-one correspondence with elements of this space via the rule $\psi(x) = \langle x|\psi\rangle$ for any state $\psi$,[1] so quantities quadratic in $\psi(x)$ must correspond to products of elements in the Hilbert space. This requires an algebraic structure on the quantum state space, and there is no candidate. Alternatively, one can recall that as a self-respecting Hilbert space, the space of quantum states carries a positive-definite Hermitian inner product. Time evolution is implemented on states via a unitary operator of the form $\exp(-i\hbar^{-1}\hat{H}t)$, so by virtue of the Hermiticity of the inner product, it is conserved under time evolution. If terms of quadratic order or higher were allowed in the fundamental wave equation, then this desirable property would be destroyed. This is sometimes expressed to saying the the resulting theory violates unitarity. In physical terms, this corresponds to probability being created or destroyed under time evolution, which is something that would be hard to accommodate in our description of the world.[2]

In the late 1940s problems like $\beta$-decay were shown to be more or less of the trivial sort when it was discovered that a gauged quantum field theory was the correct framework for a theory of quantum electrodynamics. By the early 1970s it was understood that the weak and strong forces are also described by a quantum field theory, this time with non-Abelian symmetry generators. The full $SU(3) \times SU(2) \times U(1)$ theory described all known particle interactions, predicted a number of new particles (the $Z$ and $W^{\pm}$) whose masses and interactions were subsequently shown to be in perfect agreement with the predictions of the theory. The gauge symmetries which this theory imposed upon the Lagrangian, together with some technical ideas like unitarity gave guidance, for the first time, about what theories were acceptable to Nature. Nonetheless, all of this progress fit into the general scheme of quantization outlined in Chapter 2, in which there was no way to accommodate theories such as (3.0.15).

---

[1]Indeed in many cases the Hilbert space can be taken to be, for instance, $L^2$, in which case the wavefunctions $\psi(x)$, for a suitably restricted class of functions, are isomorphic to the quantum Hilbert space.

[2]Unitarity-violating theories are not always bad. For example, if one wishes to describe quantum mechanics in the vicinity of a black hole then one might wish to include the effects of probability being destroyed or created (in the case of a white hole) at the singularity.

The success of the quantum field theory programme made it natural to try and incorporate gravity in the same way (hoping that non-linearities and other perplexities might be smoothed away in the process), but, so far, it has proved entirely impossible to devise a consistent quantum field theory of gravity. There have been many proposals within the context of field theory for mechanisms to circumvent the difficulties and conundrums which appear, but none have been satisfactory. In recent years the balance of opinion has shifted from considering this a disaster of physics to the realization that at some level the Standard Model – verified to exquisite experimental precision as it is – may itself only be approximation or low-energy realization of a deeper theory. From this point of view, the only distinction between gravity and the other forces (electroweak and strong) is that the strong/electroweak theory can be approximated at low energy by a well behaved field theory, whereas gravity can not. About the nature of whatever fundamental theory actually exists there is no firm evidence, and only a few ambiguous clues: to date, the best evidence for physics beyond the standard model comes from observations at the Sudbury Neutrino Observatory of non-zero neutrino mass, and possibly measurements of CP violation, which are both without explanation in the Standard Model. On the other hand, there is much stronger evidence from cosmology, which requires the introduction of a dark matter species (or several dark matter species) and a microphysical mechanism to produce a vacuum energy density. Cosmology itself cannot provide us with detailed particle physics for these phenomena at the level of accelerator experiments, but one can indirectly hope to gather information about masses, couplings, and decay rates.

## 3.1. The Polyakov action and the string spectrum

The most promising of the current beyond-the-Standard-Model theories is string theory.

String theory is the theory of a two-dimensional quantum field theory propagating on a two-dimensional manifold or sheet $M$ (referred to as the world sheet) injected into a background spacetime. Let $g_{ab}$ be a Euclidean metric on the worldsheet with coordinates $\sigma^a$, and let $X^\mu(\sigma)$ be the embedding of the worldsheet into spacetime. The Polyakov action for this theory is

$$\int [\mathrm{d}X][\mathrm{d}g] \exp\left(-\frac{1}{4\pi\alpha'}\int_M \mathrm{d}^2\sigma\ \sqrt{g}g^{ab}\partial_a X^\mu \partial_b X_\mu + \frac{\lambda}{4\pi}\int_M \mathrm{d}^2\sigma\ \sqrt{g}R + \frac{\lambda}{2\pi}\int_{\partial M} \mathrm{d}s\ K\right) \tag{3.1.1}$$

where $g = \det g_{ab}$, $R$ is the Ricci curvature of the worldsheet, and $K$ is the trace of the second fundamental form on the worldsheet boundary $\partial M$. The measure $\mathrm{d}s$ describes proper distance along the boundary. The term involving the embedding $X^\mu$ is called the matter theory, and can be replaced with any arbitrary two-dimensional conformal field theory, subject to some technical restrictions which will be described below.[3] The coupling $\alpha'$ is called the Regge slope and determines the string tension.

The action (3.1.1) is trivially invariant under diffeomorphisms of $\sigma^a$, because it was written in covariant form. A less trivial invariance appears when one subjects the worldsheet metric $g_{ab}$ to a local Weyl (or conformal) transformation,

$$g_{ab} \mapsto g'_{ab} = \mathrm{e}^{2\omega(\sigma)} g_{ab}. \tag{3.1.3}$$

This is just a local rescaling of the metric. A theory invariant under such a transformation is called conformal, or a conformal field theory. (The abbreviation CFT is usually used for convenience.) In this case, the invariance under Weyl transformations is guaranteed because we are working on a two-dimensional worldsheet.

As was outlined in Chapter 2, the path integral (3.1.1) is formally infinite as it stands because of divergent integrations over the gauge group diff $\times$ Weyl. To render it finite one employs the diffeomorphsm and Weyl invariance to bring $g_{ab}$ to a fiducial form $\hat{g}_{ab}$, for which one usually uses the conformal gauge, $\hat{g}_{ab} = \mathrm{e}^{2\omega(\sigma)}\delta_{ab}$. The integration over $[\mathrm{d}g]$ in the Polyakov functional integral can then be eliminated. The gauge-fixed functional integral is[4]

$$Z = \int [\mathrm{d}X]\, \Delta_{\mathrm{FP}}(\hat{g}) \exp\left( -\frac{1}{4\pi\alpha'} \int_M \mathrm{d}^2\sigma\, \partial^a X^\mu \partial_a X_\mu \right). \tag{3.1.4}$$

---

[3]The remaining two terms constitute a coupling $\lambda$ multiplied by the (modified) Einstein action for the worldsheet. In two dimensions, the Einstein action is a topological invariant corresponding to the Euler number, so the presence of this factor affects only the relative weighting that topologically distinct worldsheets receive in the path integral. Based on these considerations, it is fairly easy to argue (Polchinski, 1998) that $\lambda$ sets the open- and closed-string coupling constants,

$$g^2_{\mathrm{open}} \sim g_{\mathrm{closed}} \sim \mathrm{e}^{\lambda}. \tag{3.1.2}$$

Although $\lambda$ may appear to be a free parameter, its expectation value turns out to be determined by the string dynamics.

[4]neglecting the topological term for clarity, since, as described above, this only determines the open and closed string coupling.

where $\Delta_{\rm FP}(\hat{g})$ is the Fadeev–Popov determinant. (We are not using generalized $\xi$-gauges here, but instead working directly with a gauge-fixing $\delta$-functional, so no extra gauge-fixing terms need appear in the action.) The gauge-fixing functional is the tensor-valued expression $\Lambda_{ab} = g_{ab} - \hat{g}_{ab} = 0$, or

$$\Lambda_{ab} = g_{ab} - \mathrm{e}^{2\omega}\delta_{ab} = 0. \tag{3.1.5}$$

From (2.4.28), the Fadeev–Popov determinant can be calculated from a infinitesimal gauge transformation applied to the gauge fixing functional. Applying a combination of Weyl rescaling and coordinate diffeomorphisms to $g_{ab}$, the most general variation is

$$\begin{aligned} \delta\Lambda_{ab} &= 2\delta\omega\hat{g}_{ab} - \hat{\nabla}_a\delta\sigma_b - \hat{\nabla}_b\delta\sigma_a \\ &= (2\delta\omega - \hat{\nabla}_c\delta\sigma^c)\hat{g}_{ab} - 2\hat{P}_{ab|c}\delta\sigma^c \end{aligned} \tag{3.1.6}$$

where a hat over any quantity indicates, as usual, that it is built out of the gauge-fixed metric and connexion, and $\hat{P}_{ab|c}$ is a tensor-valued operator,

$$\hat{P}_{ab|c}\delta\sigma^c = \frac{1}{2}\left(\hat{\nabla}_a\delta\sigma_b + \hat{\nabla}_b\delta\sigma_a - \hat{g}_{ab}\hat{\nabla}_c\delta\sigma^c\right), \tag{3.1.7}$$

and we are adopting the notation of Polchinski (1998). This variation is parametrized by a scalar $\delta\omega$, and a vector $\delta\sigma^c$ which lie in a representation $(\delta\omega, \delta\sigma^c)$ of the gauge algebra diff $\times$ Weyl. The Fadeev–Popov operator is a differential form on this algebra which takes values in second-rank worldsheet tensors, as follows

$$\begin{aligned} \delta\Lambda_{ab} &= (\Box_{\rm FP})_{ab}\begin{pmatrix} \delta\omega \\ \\ \delta\sigma^c \end{pmatrix} \\ &= \begin{pmatrix} 2g_{ab} & -g_{ab}\hat{\nabla}_c - 2\hat{P}_{ab|c} \end{pmatrix}\begin{pmatrix} \delta\omega \\ \\ \delta\sigma^c \end{pmatrix} \end{aligned} \tag{3.1.8}$$

This can be differentiated at once to find $(\Box_{\rm FP})_{ab}$. Using the standard procedure described in Section 2.4.2 to rewrite the Fadeev–Popov determinant in terms of ghost fields, we find that

$$\begin{aligned} \text{ghost action} &\propto \int \mathrm{d}^2\sigma\,\sqrt{\hat{g}}\,b^{ab}(\Box_{\rm FP})_{ab}\begin{pmatrix} w \\ \\ c^d \end{pmatrix} \\ &= \int \mathrm{d}^2\sigma\,\sqrt{\hat{g}}\,b^{ab}\left[(2w - \hat{\nabla}_d c^d)\hat{g}_{ab} - 2\hat{P}_{ab|d}c^d\right], \end{aligned} \tag{3.1.9}$$

where $w$, $c^d$ and $b^{ab}$ are respectively an anticommuting scalar, vector, and second-rank tensor. These are the ghost fields of the theory. As always when dealing with integrals over anticommuting variables, the integrals are understood to be defined in a Berezin sense.[5] The normalization of the fields $w$, $c^d$ and $b^{ab}$ can be chosen conveniently. However, there is an immediate simplification one can effect by integrating out the scalar $w$, which produces an effective $\delta$-functional enforcing $\delta_{\mathrm{D}}(b^{ab}\hat{g}_{ab})$. This means that the ghost field $b^{ab}$ can be restricted to be traceless (in a flat metric): with this choice, the gauge-fixed Polyakov action reduces to

$$Z = \int [\mathrm{d}X][\mathrm{d}\bar{b}^{ab}][\mathrm{d}c^d]\ \exp\left(-\frac{1}{4\pi\alpha'}\int_M \mathrm{d}^2\sigma\ \hat{\nabla}_a X^\mu \hat{\nabla}^a X_\mu + \frac{1}{2\pi}\int \mathrm{d}^2\sigma\ \sqrt{\hat{g}}\ \bar{b}^{ab}\hat{P}_{ab|d}c^d\right), \tag{3.1.11}$$

where $\bar{b}^{ab}$ is traceless.

**3.1.1. The string spectrum.** In what follows we drop the decorations on quantities such as $\hat{\nabla}$ and $\bar{b}^{ab}$, and assume it is understood that the metric which appears is always the gauge fixed metric, and that $b^{ab}$ is a traceless ghost. The present discussion can be put on a rigorous level in the context of BRST quantization, as discussed in Section 2.5, but for present purposes we merely borrow ideas and notation from the BRST formalism and carry out the quanitzation in a brief (but *ad hoc*) manner. The proper BRST quantization is needed to take account of the ghost contributions, but this would take us too far afield into worldsheet techniques.

Consider any action $S = \int \mathrm{d}^2\sigma\,\mathcal{L}$, which under a general perturbation of the metric $g_{a_\mathrm{b}} \mapsto g_{ab} + \delta g_{ab}$ has variation

$$\delta S = \int \mathrm{d}^2\sigma\ \frac{\delta\mathcal{L}}{\delta g_{ab}}\delta g_{ab} = -\int \mathrm{d}^2\sigma \frac{\sqrt{g}}{4\pi}T^{ab}g_{ab}, \tag{3.1.12}$$

[5] Although integrals over anticommuting variables are well-known and familiar to physicists, nomenclature is sometimes not standardized. Anticommuting integration of the sort we are calling Berezin integration is also referred to as Grassmann integration in some textbooks. Up to sign, there is only one consistent choice,

$$\int \mathrm{d}\psi = 0 \quad \text{and} \quad \int \mathrm{d}\psi\ \psi = 1. \tag{3.1.10}$$

One can either define this beginning with Grassmann differentiation and choosing integration to be the formal inverse operator, or by noting that the important property of action integrals in the path integral forulation is that of translation invariance, and defining anticommuting integration so that this property is preserved (Peskin and Schroeder, 1995; Weinberg, 1994).

where $T^{ab}$ is *defined* to be the energy–momentum tensor of this action.[6] This tensor is covariantly conserved, $\nabla_a T^{ab} = 0$ as a consequence of translation invariance of the action. If $\delta g_{ab}$ is a Weyl transformation, of the form $\delta g_{ab} \propto g_{ab}$, and the action $S$ is invariant under such transformations, then the energy–momentum tensor $T^{ab}$ must satisfy $g_{ab}T^{ab} = 0$, or, more simply, $T^{ab}$ must be traceless. In fact, more is true, since under *any* gauge transformation of the metric the matrix elements of physical states (defining scattering amplitudes or S-matrix elements, as discussed for BRST symmetry above) must be invariant. Therefore,

$$\text{under } g_{ab} \mapsto g_{ab} + \delta g_{ab}, \quad \text{we have} \quad \delta\langle\alpha|\beta\rangle = \delta g_{ab}\langle\alpha|T^{ab}|\beta\rangle = 0. \tag{3.1.13}$$

If this is to be true for arbitrary physical states $|\alpha\rangle$ and $|\beta\rangle$, then one must have $\langle\alpha|T^{ab}|\beta\rangle = 0$. This is analogous, in the present context, to the requirement that physical states be BRST invariant. In addition, there are the expected analogues of BRST-exact states, as we now describe.

The energy–momentum tensor, ignoring ghosts, is

$$T_{ab} = -\frac{1}{\alpha'}\left(\nabla_a X^\mu \nabla_b X_\mu - \frac{1}{2} g_{ab} \nabla_c X^\mu \nabla^c X_\mu\right), \tag{3.1.14}$$

and the equation of motion constrains the $X^\mu$ to obey

$$\nabla_a \nabla^a X^\mu = 0 \quad \text{or,} \quad \frac{1}{g^{1/2}} \partial_a \left(g^{1/2} g^{ab} \partial_b X^\mu\right) = 0. \tag{3.1.15}$$

We are dealing with the Euclidean theory with coordinates $\{\sigma^1, \sigma^2\}$, which relate to the coordinates of the Lorentz theory via $\sigma^1 \mapsto \sigma$ and $\sigma^2 \mapsto \mathrm{i}\tau$. The Euclidean time $\sigma^2$ runs from $-\infty$ to $\infty$, but the spatial coordinate $\sigma^1$ requires more care. It cannot stretch to $\infty$, since this would represent the unphysical[7] case of an infinitely long string. Instead, $\sigma^1$ must be bounded or periodic. The bounded case corresponds to open strings; the periodic case corresponds to closed strings.

Consider closed strings. Choose coordinates so that the periodicity is $\sigma^1 \sim \sigma^1 + 2\pi$, and define complex coordinates $w$ and $z$ on the worldsheet via

$$w = \sigma^1 + \mathrm{i}\sigma^2 \quad \text{and} \quad z = \mathrm{e}^{-\mathrm{i}w}. \tag{3.1.16}$$

[6] As conventional in string theory, this definition of the energy–momentum tensor contains an extra factor of $-(2\pi)^{-1}$ in comparison with the usual convention in general relativity (Polchinski, 1998).

[7] The energy cost of stretching an infinite string with non-zero tension is, clearly, infinite.

This is a natural construction in two dimensions, where the worldsheet can be identified as a Riemann surface. The coordinate transformation which inverts $z$ is

$$2\sigma_2 = \ln \bar{z} + \ln z \quad \text{and} \quad 2\mathrm{i}\sigma_2 = \ln \bar{z} - \ln z, \tag{3.1.17}$$

where $\bar{z}$ is the complex conjugate of $z$. The metric, in unit gauge, in terms of $z$ and $\bar{z}$ is

$$\mathrm{d}s^2 = \frac{\mathrm{d}z\,\mathrm{d}\bar{z}}{|z|^2}, \tag{3.1.18}$$

where $|z|^2 = z\bar{z}$ and it is conventional to write the derivatives $\nabla_z$, $\nabla_{\bar{z}}$ as $\partial$, $\bar{\partial}$ respectively. We have already noticed above that the Polyakov action is invariant under Weyl transformations, and therefore that the energy–momentum tensor is traceless, $g^{ab}T_{ab} = 0$. In terms of $z$, tracelessness of $T_{ab}$ has a simpler expression:

$$g^{ab}T_{ab} = 4|z|^2 T_{z\bar{z}} = 0 \quad \text{so,} \quad T_{z\bar{z}} = 0. \tag{3.1.19}$$

The off-diagonal elements of $T_{ab}$ vanish. Conservation of energy now reduces to two simple statements,

$$\nabla^a T_{ab} = 0 \quad \text{requires} \quad \begin{cases} \bar{\partial} T_{zb} = 0 \\ \partial T_{\bar{z}b} = 0 \end{cases} \tag{3.1.20}$$

However, since there are no off-diagonal elements, this is the same as demanding that $T_{zz}$ and $T_{\bar{z}\bar{z}}$ are holomorphic and anti-holomorphic functions of $z$, respectively,

$$\begin{aligned} \bar{\partial} T_{zz} = 0 \quad &\text{implies} \quad T_{zz} = T(z) \\ \partial T_{\bar{z}\bar{z}} = 0 \quad &\text{implies} \quad T_{\bar{z}\bar{z}} = \tilde{T}(\bar{z}). \end{aligned} \tag{3.1.21}$$

In order to minise clutter in equations, it is very convenient to work with the functions $T$, $\tilde{T}$ rather than $T_{zz}$, $T_{\bar{z}\bar{z}}$. Since $T$ and $\tilde{T}$ are holomorphic (respectively, anti-holomorphic), they have Laurent expansions in terms of $z$ (respectively, $\bar{z}$),

$$T(z) = \sum_{m=-\infty}^{\infty} \frac{L_m}{z^{m+2}} \quad \text{and} \quad \tilde{T}(z) = \sum_{m=-\infty}^{\infty} \frac{\tilde{L}_m}{\bar{z}^{m+2}}. \tag{3.1.22}$$

The Laurent coefficients $L_m$, $\tilde{L}_m$ are called the Virasoro generators. From the familiar rules of complex analysis, they can be written as contour integrals over $T(z)$ or $\tilde{T}(\bar{z})$,

$$L_m = \oint_{\mathcal{C}} \frac{\mathrm{d}z}{2\pi\mathrm{i}z} z^{m+2} T(z) \quad \text{and} \quad \tilde{L}_m = \oint_{\mathcal{C}} \frac{\mathrm{d}\bar{z}}{2\pi\mathrm{i}\bar{z}} \bar{z}^{m+2} \tilde{T}(\bar{z}) \tag{3.1.23}$$

where $\mathcal{C}$ is any contour encircling the origin anticlockwise. For the open string, there is only one set of Virasoro generators $L_m$ (Polchinski, 1998).

On quantization, the components $T(z)$ and $\tilde{T}(\bar{z})$ (and therefore the Virasoro generators $L_m$, $\tilde{L}_m$) are promoted to operators. Enforcing $T_{ab} = 0$, in terms of the Laurent coefficients, is equivalent to $L_m = 0$. Since the energy–momentum tensor is Hermitian (as a well defined quantum-mechanical operator), the Laurent coefficients must obey $L_m^\dagger = L_{-m}$ and therefore one can deal only with the $m \geqslant 0$ operators. Physical states are defined by

$$(L_m + a\delta_{m,0})|\alpha\rangle = 0 \quad \text{for } m \geqslant 0 \tag{3.1.24}$$

where one must allow for the possibility of an operator ordering constant at $m = 0$ (Green et al., 1987; Polchinski, 1998). As we mentioned above, there are also analogues of BRST-exact states, which can be constructed by operating with $m < 0$ Virasoro generators. Such states have the general form

$$|\beta\rangle = \sum_{m=1}^{\infty} L_{-m}|\beta_m\rangle \tag{3.1.25}$$

where any number of the $|\beta_n\rangle$ may be zero. States of the form $|\beta\rangle$ are orthogonal to all physical states, and if physical themselves are called null. The physical Hilbert space of the theory is essentially the cohomology of the Virasoro generators,

$$\text{Hilbert space} = \frac{\text{physical states}}{\text{null states}}. \tag{3.1.26}$$

The full details, while not complicated, are lengthy, so we omit calculations and present only the conclusions. On writing the equation of motion $\nabla^a \nabla_a X^\mu = 0$ in complex $z$-coordinates, one obtains $\partial\bar{\partial}X^\mu = 0$. Thus, $\bar{\partial}X^\mu$ is an anti-holomorphic function, whereas on taking complex conjugates, $\bar{X}^\mu$ is holomorphic. Making use of Laurent expansions once more gives

$$\partial X^\mu = -\mathrm{i}\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{m=-\infty}^{\infty} \frac{\alpha_m^\mu}{z^{m+1}} \quad \text{and} \quad \bar{\partial}X^\mu = -\mathrm{i}\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{m=-\infty}^{\infty} \frac{\tilde{\alpha}_m^\mu}{\bar{z}^{m+1}}, \tag{3.1.27}$$

where the overall normalization is conventional. One can write the Laurent coefficients $\alpha^\mu$, $\tilde{\alpha}^\mu$ more explicitly,

$$\alpha_m^\mu = \left(\frac{2}{\alpha'}\right)^{1/2} \oint_{\mathcal{C}} \frac{\mathrm{d}z}{2\pi}\, z^m \partial X^\mu(z) \quad \text{and} \quad \tilde{\alpha}_m^\mu = \left(\frac{2}{\alpha'}\right)^{1/2} \oint_{\mathcal{C}} \frac{\mathrm{d}\bar{z}}{2\pi}\, \bar{z}^m \bar{\partial} X^\mu(\bar{z}). \tag{3.1.28}$$

In the case $m = 0$ these expressions both reduce to $\oint \mathrm{d}X^\mu$, so $\alpha_0^\mu$ and $\tilde{\alpha}_0^\mu$ must be equal. Moreover, these oscillators have an important spacetime interpretation as carrying the

bulk momentum of the string, which follows just from the fact that they are the zero-modes of the string, or alternatively from the Noether current for spacetime translations, $j_a^\mu = \mathrm{i}\partial_a X^\mu/\alpha'$, the conserved charge associated with which is

$$p^\mu = \frac{1}{2\pi\mathrm{i}}\oint_{\mathcal{C}}(\mathrm{d}z\, j^\mu - \mathrm{d}\bar{z}\,\bar{\jmath}^\mu) = \left(\frac{2}{\alpha'}\right)^{1/2}\alpha_0^\mu = \left(\frac{2}{\alpha'}\right)^{1/2}\tilde{\alpha}_0^\mu. \tag{3.1.29}$$

This is just Green's theorem in the plane. The explicit mode expansions (3.1.27) can now be integrated to give a direct expression for $X^\mu$ itself, in terms of the oscillator modes $\alpha_m^\mu$, $\tilde{\alpha}_m^\mu$, $X^\mu(z,\bar{z}) = X(z) + X(\bar{z})$,

$$X^\mu(z,\bar{z}) = x^\mu - \mathrm{i}\frac{\alpha'}{2}p^\mu \ln|z|^2 + \mathrm{i}\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\frac{1}{m}\left(\frac{\alpha_m^\mu}{z^m} + \frac{\tilde{\alpha}_m^\mu}{\bar{z}^m}\right), \tag{3.1.30}$$

where $x^\mu$ is a constant of integration that describes the string's centre of mass trajectory, in the same way that $p^\mu$ describes its bulk momentum. When quantized, the canonical commutation relations imply

$$[\alpha_m^\mu, \alpha_n^\nu] = [\tilde{\alpha}_m^\mu, \tilde{\alpha}_n^\nu] = m\delta_{m,-n}\eta^{\mu\nu} \quad \text{and} \quad [x^\mu, p^\nu] = \mathrm{i}\eta^{\mu\nu}, \tag{3.1.31}$$

so the oscillators $\alpha_m^\mu$, $\tilde{\alpha}_m^\mu$ act as raising and lowering operators on the string quantum state, up to a constant of proportionality. In particular, the $m > 0$ modes are lowering operators – analogous to the annihilation operators $a$ of field theory – whereas the $m < 0$ modes are raising operators, analogous to the creation operator $a^\dagger$. The mode number $m$ labels the oscillator for which $\alpha_m^\mu$ and $\tilde{\alpha}_m^\mu$ are annihilation and creation operators, just as the Minkowski space operators $a(\mathbf{k})$ and $a^\dagger(\mathbf{k})$ destroy and annihilate particles of momentum $\mathbf{k}$. The Laurent coefficients can be written as

$$L_m \sim \frac{1}{2}\sum_{n=-\infty}^{\infty}\alpha_{m-n}^\mu\alpha_{\mu,n} \quad \text{and} \quad \tilde{L}_m \sim \frac{1}{2}\sum_{n=-\infty}^{\infty}\tilde{\alpha}_{m-n}^\mu\tilde{\alpha}_{\mu,n}, \tag{3.1.32}$$

up to operator ordering concerns.

We can now state the string spectrum.

(1) Open strings. Begin with the string vacuum, which is annihilated by all lowering operators. This is not the spacetime vacuum $|\Omega\rangle$, but instead labels a state of the string which is not excited. This state is written $|0;k\rangle$ and may carry a non-zero momentum $k^\mu$ corresponding to the bulk momentum of the string.

To impose the physical state conditions, one needs the Virasoro generators (3.1.23). The only generator which does not automatically annihilate the vacuum

is the zero-generator $L_0$ (or $\tilde{L}_0$)

$$L_0 = \alpha' p^2 + \alpha_{-1} \cdot \alpha_1 + \cdots , \tag{3.1.33}$$

so bearing in mind the possible operator-ordering constant (3.1.24), this is a condition on the spacetime momentum,

$$\alpha' k^2 + a = 0 \quad \text{or, since } k^2 = -m^2, \quad m^2 = \frac{a}{\alpha'}, \tag{3.1.34}$$

where $m^2$ is the spacetime mass, or the mass of the string as seen from on observer in spacetime. The inner product $\cdot$ is taken in the spacetime metric, consisting of contractions of Greek indices.

The first excited level is obtained by operating on the vacuum with the creation operator $\alpha^{\mu}_{-1}$,

$$|e; k\rangle = e \cdot \alpha_{-1} |0; k\rangle, \tag{3.1.35}$$

in which we have introduced a polarization tensor $e^{\mu}$. The norm is

$$\langle e; k | e; k' \rangle = \langle 0; k | (\bar{e} \cdot \alpha_1)(e \cdot \alpha_{-1}) | 0; k' \rangle \propto \bar{e}^{\mu} e_{\mu} \delta_{\mathrm{D}}(k - k'), \tag{3.1.36}$$

where $\bar{e}$ is the complex conjugate of $e$, so the timelike oscillator has a negative norm. Recalling that observables are supposed to provide a map from states to probability measures, negative norm states cannot be permitted under the normal rules of quantum mechanics, since they would appear as states of negative probability. This catastrophe is avoided by the physical state conditions, which remove the pathological state of negative norm, leaving a healthy, unitary quantum theory. The $L_0$ constraint supplies a mass condition,

$$L_0 |e; k\rangle = (-m^2 \alpha^2 + 1) |e; k\rangle, \tag{3.1.37}$$

after commuting the $\alpha_{\pm 1}$ oscillators, whereas the only term in $L_{\pm 1}$ which does not annihilate $|e; k\rangle$ is

$$L_{\pm 1} = (2\alpha')^{1/2} p \cdot \alpha_{\pm 1} + \cdots \tag{3.1.38}$$

Applying $L_1$ to $|e; k\rangle$ gives

$$(p \cdot \alpha_1)(e \cdot \alpha_{-1}) |e; k\rangle = e \cdot k |e; k\rangle = 0. \tag{3.1.39}$$

Therefore the polarization $e^\mu$ should be orthogonal to the momentum $k_\mu$. Moreover the excited state at this level which is generated by $L_{-1}$ is spurious,

$$L_{-1}|0;k\rangle = (2\alpha')^{1/2} k \cdot \alpha_{-1}|0;k\rangle; \tag{3.1.40}$$

so the choice $e^a \propto k^a$ is spurious. There are three choices,

(a) The operator ordering constant $a$ satisfies $a > -1$. In this case, the mass-squared for the excitation is positive. Moreover, in the rest frame of the string, the momentum is $k = (m, \mathbf{0})$ say, with $m$ real, so the physical state condition removes the timelike polarization. The spurious state is not physical, because $k \cdot k \neq 0$. There are no negative norm states. This excitation corresponds to a massive vector boson.

(b) $a = -1$. The mass of the excitation is zero. The momentum must become null, so the spurious state becomes physical. There are no negative norm states. This excitation corresponds to a massless vector boson.

(c) $a < -1$. Now the momentum is spacelike, so this theory already has pathologies. The spurious state is not physical. This is the theory of a tachyonic vector boson with negative norm states.

From gauge theory, we know that massive vector bosons are not interesting, whereas the tachyonic vector boson is pathological. Therefore we must choose this excitation to correspond to a massless vector boson $A_\mu$.

(2) Closed strings. The situation is much the same, except that there are two sets of oscillators $\alpha_m^\mu$ and $\tilde{\alpha}_m^\mu$. The vacuum is

$$|0;k\rangle \quad \text{with mass} \quad m^2 = -\frac{4}{\alpha'} \tag{3.1.41}$$

and the first excited state is

$$e_{\mu\nu}\alpha_{-1}^\mu \tilde{\alpha}_{-1}^\nu |0;k\rangle \quad \text{where} \quad m^2 = 0, k^\mu e_{\mu\nu} = k^\nu e_{\mu\nu} = 0, \tag{3.1.42}$$

where we have set the operator ordering constant to $-1$. This corresponds to a spin-2 excitation like the graviton perturbation $h_{ab}$, but one which is not necessarily symmetric. Therefore $e_{\mu\nu}$ must decompose into a antisymmetric part $B_{\mu\nu}$, and a symmetric part which can be further reduced to a traceless symmetric tensor $G_{\mu\nu}$ and an overall scalar mode $\phi$ which encodes the trace.

The tachyons are an artefact of the fact that this theory is not supersymmetric,[8] and disappear when fermionic degrees of freedom are coupled to the theory.

**3.1.2. Strings in background fields.** At low energy, only the massless modes are important. These are the massless vector $A_\mu$ from the open string sector, and the graviton $G_{\mu\nu}$, the antisymmetric tensor (or Neveu–Schwarz two-form) $B_{\mu\nu}$ and the scalar (or dilaton) $\phi$. More massive modes have characteristic masses of order the Planck scale, and decouple from the low energy world.

If string theory really is an ultra-violet completion of general relativity, then one should expect that the dynamics of at least the massless graviton mode $G_{\mu\nu}$ could be written as general relativity with extra material contributions from the matter fields $A_\mu$, $B_{\mu\nu}$ and $\phi$. However, $G_{\mu\nu}$ is a perturbative graviton, corresponding to excitation $h_{ab}$ in $\mathrm{d}s^2 = (g_{ab} + h_{ab})\mathrm{d}x^a\mathrm{d}x^b$. In order to study backgrounds which are more than perturbative departures from Minkowski space, it is neccessary to introduce condensations of gravitons, corresponding to a non-trivial background spatial metric, and possibly non-trivial background fields for $A_\mu$, $B_{\mu\nu}$ and $\phi$ also.

A generalized worldsheet metric which couples to non-trivial backgrounds is supplied by the modified Polyakov action,

$$S_{\mathrm{P}} = \frac{1}{4\pi\alpha'}\int \mathrm{d}^2\sigma\ \sqrt{g}\left[(g^{ab}G_{\mu\nu} + \mathrm{i}\varepsilon^{ab}B_{\mu\nu})\nabla_a X^\mu \nabla_b X^\nu + \alpha' R\phi(X)\right]. \tag{3.1.43}$$

Demanding Weyl invariance of this action yields the following equations of motion for $G_{\mu\nu}$, $B_{\mu\nu}$ and $\phi$, up to terms of order $\alpha'^2$,

$$\begin{aligned} \alpha' R^G_{\mu\nu} + 2\alpha'\nabla_\mu\nabla_\nu\phi - \frac{\alpha'}{4}H_{\mu\lambda\omega}H_\nu{}^{\lambda\omega} + \cdots = 0 \\ -\frac{\alpha'}{2}\nabla^\omega H_{\omega\mu\nu} + \alpha'\nabla^\omega\phi H_{\omega\mu\nu} + \cdots = 0 \\ \frac{D-26}{6} - \frac{\alpha'}{2}\triangle\phi + \alpha'\nabla_\omega\phi\nabla^\omega\phi - \frac{\alpha'}{24}H_{\mu\nu\lambda}H^{\mu\nu\lambda} + \cdots = 0, \end{aligned} \tag{3.1.44}$$

[8]The open string tachyon has an interesting interpretation in terms of D-brane annihilation, and is potentially of interest in cosmology via the so-called Chaplygin gas. This is a gas obeying an equation state of the form $p = k/\rho$ for some constant $k$, which can mimic some aspects of the cosmological consequences of dark matter or vacuum energy (Barreiro and Sen, 2004; Bento, Bertolami, and Sen, 2003; Sen, 1998, 1999, 2002a,b, 2003).

At the present time, it seems that there is no known corresponding interpretation of the closed string tachyon.

where $R^G$ is the curvature built from $G_{\mu\nu}$, not the worldsheet curvature which is built from $g_{ab}$, and $H_{\omega\mu\nu}$ is the field strength

$$H_{\omega\mu\nu} = \nabla_\omega B_{\mu\nu} + \nabla_\mu B_{\nu\omega} + \nabla_\nu B_{\omega\mu}. \tag{3.1.45}$$

These equations of motion can be derived from the spacetime action principle

$$S = \frac{1}{2\kappa^2} \int \mathrm{d}^D x \, \sqrt{-G} \mathrm{e}^{-2\phi} \left[ -\frac{2(D-26)}{3\alpha'} + R^G - \frac{1}{12} H_{\mu\nu\lambda} H^{\mu\nu\lambda} + 4\nabla_\mu \phi \nabla^\mu \phi + O(\alpha') \right]. \tag{3.1.46}$$

This is called the sigma model action, and describes the low energy bosonic excitations of string theory.

## 3.2. Compactification to low dimensions

In this section, we wish to outline the basic features of string compactifications. Ideas from this technology will be applied in Part 2 together with elementary D-branes, to be described in Section 3.3 below, to discuss string physics in an cosmological setting.

The basic results we will need are the Kaluza–Klein mechanism, and a description of how zero modes arise in the various four-dimensional fields, so there is no need to discuss the detailed and technical topological theorems which are usually implicit in a discussion of string compactifications (Green et al., 1987). In particular we are not going to discuss fermions, so there is no need for an extensive excursion to deal with the Dirac index. Instead, most of the technology is cohomological and was already outlined in Section B.3.1. A fuller outline of the whole subject, based largely on topological index theorems, can be found in Chapters 14–15 of Green et al. (1987), or Greene (1997).

In almost all of the foregoing analysis, we left the dimension of spacetime unspecified. It turns out that quantum string theory of the kind we have described does not make sense on spacetimes of all dimensionalities, but only for a specific dimension in which the quantum theory retains the Weyl invariance which was exploited to put the Polyakov action in the particularly simple gauge-fixed form (3.1.11). At one level this entails a fairly complicated procedure[9] in which one must carefully calculate the regularized trace of the energy–momentum tensor and ensure that it remains zero. However, it is easy to see intuitively that the quantum theory must remain Weyl invariant, otherwise it will depend on the choice of the gauge-fixed metric $\hat{g}_{ab}$, which is unphysical.

[9] Although a simplified argument can be given in light cone gauge; see, for example, Polchinski (1998).

The condition that Weyl invariance is preserved is equivalent, in the bosonic string theory discussed above, to the requirement that the dimensionality $d$ of spacetime satisfy $d = 26$. This is rather large in comparison with the four large dimensions we see in the world around us, but one can remove a few of these dimensions by working with the full supersymmetric string theory, which contains fermions $\psi^\mu$ in addition to the bosonic fields $X^\mu$. These theories obey a spacetime supersymmetry,[10] and have fewer dimensions, needing only ten spacetime dimensions to be consistent. Although we are working only in the bosonic sector, one should really frame all discussion in terms of the supersymmetric version. Apart from any other considerations, these string theories have perturbatively well-behaved vacua, without the appearance of tachyonic matter. Therefore in the following discussion, and throughout this thesis in general, we assume that one wishes to compactify a ten-dimensional string theory to a low-dimensional universe.

In this section, the low energy world is four dimensional and M-theory is not named. Almost all compactification technology survives the transition to M-theory, so it is only necessary to point out a few extra features when discussing five-dimensional compactifications.

**3.2.1. The wave operator and the Kaluza–Klein mechanism.** Consider some general ten dimension stringy background $M$, which is taken to be of the form $M_4 \times K$, where $M_4$ is four-dimensional Minkowski space, or some other four-dimensional space of interest, and $K$ is a compact 6-dimensional manifold called the compactification manifold. The analysis involves the application of Hodge–de Rham theory as outlined in Section B.2.1.1 and Section B.3.5, and Cartan structural equations described in Section B.5.2.

To simplify the description, and also to outline a case which will specifically become interesting in Part 2, consider the case with $K$ only one-dimensional, and constituting a topological circle. Fields on $M_4 \times K$ are classified, as usual in quantum theory, according to the representation of Lorentz group to which they belong, or more generally according to the tangent space group. In a $d$-dimensional Lorentzian theory, this is $SO(1, d-1)$ but may be $SO(d)$ in a Euclidean theory. A typical field of spin $s$ on $M_4 \times K$ may appear

[10] There are two distinct constructions of the superstring, in one of which (the Green–Schwarz superstring) spacetime supersymmetry is manifest, and in the other (the Ramond–Neveu–Schwarz superstring) it is not. The Green–Schwarz construction is described in text-book form only in the older monograph Green et al. (1987).

as a collection of fields on $M_4$. For example, a rank two symmetric tensor theory $g^{(5)}_{ab}$ on $M_4 \times \mathbf{S}$ decomposes into a rank two symmetric tensor $g^{(4)}_{\mu\nu}$, a vector $A_\mu$ and a scalar $\phi$ on $M_4$,

$$g^{(5)}_{ab} = \begin{pmatrix} \phi & A_\mu \\ A_\mu & g^{(4)}_{\mu\nu} \end{pmatrix}. \tag{3.2.1}$$

This pattern is repeated for representations of arbitrary spin. The field content as viewed from $M_4$ can always be found by decomposing the representation of $SO(1, d-1)$ into a representation of $SO(1,3) \times SO(6)$ (or $SO(1,3) \times SO(1)$ in the present case where $K$ is one-dimensional).

Consider a general $p$-form on $M_4 \times K$. In (B.5.7), the curvature of a connexion $\omega$ was defined to be $\Omega = \mathcal{D}\omega$, where $\mathcal{D}$ is the gauge-covariant exterior derivative. This generalizes quite straightforwardly to $p$-forms, of which the connexion $\omega$ is a special example of a 1-form, so that for a general $p$-form $B$ the field strength (or curvature) is

$$F = \mathrm{d}B \tag{3.2.2}$$

and the field equation following from the Yang–Mills Lagrangian (B.5.14) is $\delta \mathrm{d}B = 0$. We are supposing that $B$ is a scalar-valued $p$-form, so there is no need to use the gauge derivative $\mathcal{D}$ and one can employ d instead. The field strength $F$ is invariant under transformations

$$B \mapsto B' = B + \mathrm{d}\Lambda \tag{3.2.3}$$

for some $(p-1)$-form $\Lambda$. One can fix the gauge invariance assocatied with this transformation by choosing the analogue of Lorentz gauge, $\delta B = 0$. Then the field equation is equivalent to

$$\triangle B = (\triangle_{M_4} + \triangle_K) B = 0, \tag{3.2.4}$$

where $\triangle = \not{d}^2$ is the Laplacian on $M_4 \times K$, which decomposes into the sum $\triangle = \triangle_{M_4} + \triangle_K$ (Green et al., 1987).

This implies that $\triangle_K$ behaves as a mass operator for the four-dimensional theory, in which the allowed values of $\triangle_K$ consitute possible masses for the $M_4$ theory. If $B$ is separable, of the form $B = \alpha \wedge \beta$, where $\alpha$ is an $n$-form on $M_4$ and $\beta$ is a $(p-n)$-form on $K$, then

$$(\triangle_{M_4} \alpha) \wedge \beta + \alpha \wedge (\triangle_K \beta) = 0. \tag{3.2.5a}$$

If $\beta$ is an eigenform of the Laplacian, $\triangle_K \beta = -m^2 \beta$, then the $M_4$ field equation reads

$$(\triangle_{M_4} - m^2)\alpha = 0. \tag{3.2.5b}$$

In particular, the number of zero eigenvalues of $(p-n)$-forms on $K$ for $\triangle_K$ determines the number of massless $n$ forms which will appear in the four-dimensional theory on $M_4$. There are no extra combinations arising from $\triangle_{M_4}\alpha = 0$, because $M_4$ is supposed to Minkowski space, which has trivial cohomology. However, if the four-dimensional compactification manifold is topologically non-trivial then it is possible that one may acquire more massless zero modes.

It only remains to count the number of zero eigenvalues of $\triangle_K$, but this is quite straightforward in virtue of the Hodge decomposition theorem state in Section B.3.5. Each harmonic form $\triangle_K \beta_n$ is associated with a cohomology class of $K$, and the number of such classes is the Betti number $b_{p-n}$.

## 3.3. T-duality and strings at strong coupling

So far, we have been discussing low energy field theories – truncated infra-red approximations to the full string theory – on compact backgrounds. The next step up in sophistication is to consider strings on compact backgrounds. Apart from actually being useful to the material developed in Part 2, the study of open strings on a circle is the simplest route to the discovery of D-branes, which are fundamentally the objects around which this thesis is based.

**3.3.1. Closed strings.** Consider closed string theory on some geometrical background with one dimension compactified on a circle of radius $R$, so that $X \simeq X + 2\pi R$ in the compact direction. For closed strings where the spatial dimension of the worldsheet is periodic, so that $\sigma + 2\pi \simeq \sigma$, this opens up the possibility that $X$ may close only up to multiple wrappings of the compact spacetime dimension:

$$X(\sigma + 2\pi) = X(\sigma) + 2\pi R w \quad \text{where} \quad w \in \mathbf{Z}, \tag{3.3.1}$$

where $w$ is an integer called the winding number. This is a new effect peculiar to string theory, since although one can have topologically non-trivial modes in field theory, there is no analogue of the winding number. For this reason one expects new physics to be associated with $w$, as we now describe.

Let us return to the mode expansions (3.1.27). Integrating $X$ along the worldsheet gives back the winding number,

$$\oint (\mathrm{d}z\,\partial X + \mathrm{d}\bar{z}\,\bar{\partial} X) = 2\pi \left(\frac{\alpha'}{2}\right)^{1/2} (\alpha_0 - \tilde{\alpha}_0) = 2\pi R w, \tag{3.3.2}$$

and the bulk momentum of the string remains as before, (3.1.29),

$$p = \frac{1}{2\pi\alpha'} \oint (\mathrm{d}z\,\partial X - \mathrm{d}\bar{z}\,\bar{\partial} X) = \frac{1}{\sqrt{2\alpha'}} (\alpha_0 + \tilde{\alpha}_0). \tag{3.3.3}$$

This lets us identify left-moving and right-moving momenta, according to the following scheme, since the bulk momentum $k$ must obey the field theory quantization rule $k = n/R$, for integer $n$,

$$\begin{aligned} p_L &= \sqrt{\frac{2}{\alpha'}}\alpha_0 = \frac{n}{R} + \frac{wR}{\alpha'} \\ p_R &= \sqrt{\frac{2}{\alpha'}}\tilde{\alpha}_0 = \frac{n}{R} - \frac{wR}{\alpha'}. \end{aligned} \tag{3.3.4}$$

So far, this has all been for a single compact dimension. Restoring the remaining uncompactified dimensions means that the Virasoro generators should be written

$$\begin{aligned} L_0 &= \frac{\alpha'}{4}p^2 + \frac{\alpha' p_L^2}{4} + \sum_{n=1}^{\infty} \alpha_{-n}\alpha_n \\ \tilde{L}_0 &= \frac{\alpha'}{4}p^2 + \frac{\alpha' p_R^2}{4} + \sum_{n=1}^{\infty} \tilde{\alpha}_{-n}\tilde{\alpha}_n, \end{aligned} \tag{3.3.5}$$

in which $p^2$ is the momentum operator in the non-compact dimensions. These formulae just arise from splitting the compact dimension off from the remaining $p_\mu$ in (3.1.32). The mass-shell condition, obtained by imposing $L_0 = 0$ as an operator equation to give physical states, must then comprise

$$\begin{aligned} 0 &= -\frac{\alpha'}{4}m^2 + \frac{\alpha'}{4}p_L^2 + \cdots \\ 0 &= -\frac{\alpha'}{4}m^2 + \frac{\alpha'}{4}p_R^2 + \cdots \end{aligned} \tag{3.3.6}$$

where we have ignored operators giving information about the contribution of excited modes to the mass, and $m^2$ is the $(d-1)$-dimensional mass-squared in the non-compact dimensions. Collecting terms shows that

$$m^2 = \frac{n^2}{R^2} + \frac{w^2 R^2}{\alpha'^2} + \cdots . \tag{3.3.7}$$

This formula for the $(d-1)$-dimensional mass is very important. It shows that in addition to the tower Kaluza–Klein modes, $m^2 = n^2/R^2$, there is a secondary tower of winding modes. As one takes the compactification radius $R$ to be very large, $R \to \infty$, the Kaluza–Klein modes (sometimes called momentum modes) become light, whereas the the winding modes because increasingly heavy. On the other hand, if one contracts the compact dimension away by sending $R \to 0$, then the momentum modes become heavy but the winding modes become light.

In field theory the winding sector is absent, so as one unrolls the circle by sending $R \to \infty$, one recovers the field theory spectrum in $d$ uncompact dimensions. Removing the compact dimension via $R \to 0$ makes all momentum modes arbitrarily heavy, leaving only any zero modes (which are massless at any scale for the extra dimension) which constitute the field theory spectrum in $(d-1)$-dimensions. This is a dimensionally reduced theory. In string theory, the opposite happens. Although one can contract away the extra dimension, one does not end up with a dimensionally reduced theory; the winding modes become light and rebuild a $d$-dimensional field theory spectrum. Instead of removing a dimension, the contraction procedure appears to have opened up a new, infinite dimension, leaving us with a $d$-dimensional theory.

This idea can be formalized. The mass-shell condition (3.3.7) is invariant under the exchange

$$\mathrm{T} : n \leftrightarrow w, \quad R \leftrightarrow \frac{\alpha'}{R}. \tag{3.3.8}$$

This exchange is called T-duality, and swaps the momentum and winding sectors while inverting the size of circle: string theory on a circle of radius $R$ is said to be T-dual to string theory on a circle of radius $\alpha'/R$. From this observation, it is easy to see that the limits $R \to \infty$ and $R \to 0$ are actually the same, up to T-dualising. Strings view geometry at very small scales differently, with the result that the 'smallest' meaningful radius is the self-dual limit $R^2 = \alpha'$. Although we have discussed the phenomenon here only for unexcited modes and a circular compact dimension, one can show (Johnson, 2003) that T-duality is an exact non-perturbative symmetry of string theory, so that all correlation functions, scattering amplitudes or other observables computed in one string theory could be equally well computer in the T-dual theory.

There is another way to formulate T duality. Exchange of $n$ and $w$ is the same as the rule (Polchinski, 1998)

$$\mathrm{T} : p_L \mapsto p_L, \quad p_R \mapsto -p_R. \tag{3.3.9}$$

In terms of the mode expansion (3.1.30), where $X(z,\bar{z}) = X(z) + X(\bar{z})$ this can be written as,

$$\mathrm{T} : X(z,\bar{z}) \mapsto X'(z,\bar{z}) = X(z) - X(\bar{z}). \tag{3.3.10}$$

Therefore the T-dual theory is the same as the original theory, but working in terms of the coordinate $X'$ instead of $X$.

**3.3.2. Open strings.** Open strings wrapped around a compact dimension are topologically indistinguishable from unwrapped strings, because there is nothing to prevent the end-points of the string moving and continuously unwrapping the dimension. For this reason, open string do not exhibit a winding sector and the $R \to 0$ limit genuinely does dimensionally reduce the open string theory. However, this is clearly problematic because open string theories necessarily contain closed strings, so the $R \to 0$ limits must be compatible.

Returning to the integrated mode expansion (3.1.30), the open string theory is solved by the mode expansion $X^\mu(z,\bar{z}) = X^\mu(z) + X^\mu(\bar{z})$, where

$$\begin{aligned} X^\mu(z) &= \frac{x^\mu}{2} + \frac{x'^\mu}{2} - \mathrm{i}\alpha' p^\mu \ln z + \mathrm{i}\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{1}{m}\frac{\alpha_m^\mu}{z^m} \\ X^\mu(\bar{z}) &= \frac{x^\mu}{2} - \frac{x'^\mu}{2} - \mathrm{i}\alpha' p^\mu \ln \bar{z} + \mathrm{i}\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{1}{m}\frac{\tilde{\alpha}_m^\mu}{\bar{z}^m} \end{aligned} \tag{3.3.11}$$

where $x'^\mu$ is a constant which cancels out of $X^\mu(z,\bar{z})$. On T-dualising one of the dimensions one works instead in terms of the coordinate $X'$,

$$\begin{aligned} X'(z,\bar{z}) &= X(z) - X(\bar{z}) \\ &= x' - \mathrm{i}\alpha' p \ln\frac{z}{\bar{z}} + \mathrm{i}\sqrt{2\alpha'} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{1}{m}\alpha_m \mathrm{e}^{-\mathrm{i}m\sigma^2} \sin n\sigma^1 \\ &= x' + 2\alpha'\frac{n}{R}\sigma^1 + \mathrm{i}\sqrt{2\alpha'} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{1}{m}\alpha_m \mathrm{e}^{-\mathrm{i}m\sigma^2} \sin n\sigma^1, \end{aligned} \tag{3.3.12}$$

where we have reverted to the Euclidean coordinates $z = \mathrm{e}^{-\mathrm{i}\sigma^1+\sigma^2}$, where $\sigma^1 \in [0, 2\pi)$ is the spatial dimension of the worldsheet. In this representation, it is clear that the endpoints of the string $X(\sigma^1 = 0)$ and $X(\sigma^1 = 2\pi)$ do not move. Instead, they are fixed to a particular plane, up to winding terms: this plane is called a D-brane, since the strings obey Dirichlet boundary conditions there.

**3.3.3. Strings at strong coupling.** T-duality is only one of a number of dualities which relate various string theories to each other, possibly with rescaled parameters such as the T-duality radius inversion $\mathrm{T} : R \mapsto R' = \alpha'/R$. More general dualities can relate different string theories to each other, in different parameter régimes. The first step in a description of such dualities is a catalogue of the known consistent string theories.

As we have already discussed, the bosonic string theory (3.1.1) is not consistent as it stands, but must be supplemented with fermionic fields $\psi^\mu$ to produce stable perturbative vacua. Once such fields have been introduced, there is a choice about the chirality of the fermions. In addition, there are other conditions the string theory must fulfill in order to remain anomaly free. One formulation of superstring theory, known the Ramond–Neveu–Schwarz or RNS superstring, is a two-dimensional supergravity propagating over the worldsheet (Green et al., 1987; Johnson, 2003; Polchinski, 1998). This approach is analytically quite simple. An alternative formulation, which lends itself to classification of possible string theories, is the Green–Schwarz superstring. In this formulation, the string action is (Green et al., 1987)

$$S_{\mathrm{P}} = -\frac{1}{4\pi\alpha'}\int \mathrm{d}^2\sigma\ \sqrt{g}g^{ab}\Pi_a \cdot \Pi_b + \text{supplementary terms} \tag{3.3.13}$$

where $\Pi^\mu_a$ is the supercovariant completion of $\partial_a X^\mu$,

$$\Pi^a_\mu = \partial_a X^\mu - \mathrm{i}\bar\theta^A\Gamma^\mu\partial_a\theta_A, \tag{3.3.14}$$

where $\theta^A$ is a two-component worldsheet spinor. Quantum mechanically, this theory only exists in ten spacetime dimensions, and when $\theta$ is a Majorana–Weyl spinor in the Dirac representation (see the **Summary of notation** on page 15). On the worldsheet, where quantities are two-dimensional, the Dirac spinor $\theta$ is two-dimensional and its upper and lower components are chiral Weyl spinors $\theta^1$ and $\theta^2$.

$$\theta^A = \begin{pmatrix} \theta^1 \\ \\ \theta^2 \end{pmatrix}. \tag{3.3.15}$$

Either $\theta^1$ and $\theta^2$ are chosen to have the same chirality, or they are chosen to have opposite chirality. We summarise the options:

- A superstring theory based on open strings is called Type I. Of course, such a theory also contains closed strings. It turns out that the only consistent choice is to have $\theta^1$ and $\theta^2$ of the same chirality. Open strings may contain gauge degrees of freedom, which we have not discussed. At the quantum level, the only consistent choice is an $SO(32)$ gauge group.
- String theories containing closed strings only are called Type II. If $\theta^1$ and $\theta^2$ have opposite chirality, then the theory is called Type IIA. For $\theta^1$ and $\theta^2$ of the same chirality, the theory is called Type IIB. There is no gauge group.
- A final choice is to set one or other of $\theta^1$ or $\theta^2$ to zero, although this was not spelt out explicitly above. This 'empties' a sector of the matter theory, which must be filled in by some other degrees of freedom. Such theories are called heterotic, and are consistent for theories of closed strings. There are two possible gauge groups, $SO(32)$ and $E_8 \times E_8$.

All these string theories are just distinguished by different choices of the matter theory for the worldsheet, and whether they are theories of open and closed strings, or closed strings only. As such, they are still perturbative theories of strings around some background, containing some coupling $g^2_{\text{open}} = g_{\text{closed}}$, and can be expected to be a good description of physics provided that the coupling does not become large. Exactly what happens to string theory at strong coupling is more difficult to unravel.

One of the most striking, and thoroughly unexpected, discoveries of work on string theory was the realization in the mid-1990s that all five of the string theories described above (Type I; Types IIA and IIB; and the heterotic $SO(32)$ and $E_8 \times E_8$ theories) are all dual to other string theories at strong coupling.

(1) Type IIB theory. This theory is self-dual at strong coupling. The fundamental strings (or F-strings) described by the Polyakov action are supplemented in the full theory by other string-like objects built from (1+1)-dimensional D-branes (D1-branes), known as D-strings. At weak coupling, F-strings are light and dominate physics, whereas D-strings are heavy and decouple. At strong coupling, the F-string becomes heavy and exchanges roles with the D-string. This is a good

example of the kind of physics which D-branes open up: without the presence of D-branes in the theory, there is nothing to exchange roles with the F-string.

(2) Type I and heterotic $SO(32)$ theory. The same argument with D-strings can be adjusted to show that the Type I theory is dual at strong coupling to the heterotic $SO(32)$ theory.

(3) Type IIA theory. In this theory, there are no D1-branes, but rather D0-branes. At strong coupling, the spectrum of D0-brane excitations exhibits a Kaluza–Klein structure, demonstrating the existence of an extra dimension in the theory which is becoming larger as the string coupling grows. The full theory, on the extra dimension, is dubbed M-theory. Whatever this theory is, it must be 11-dimensional supergravity at low energy, because this is what the string sigma model predicts. Being a closed string theory, there is no matter, so this is a theory of supergravity alone.

(4) Heterotic $E_8 \times E_8$ theory. This leaves only the $E_8 \times E_8$ heterotic theory without a partner. In this case the techniques required to construct the strong-coupling dual are fairly refined (Horava and Witten, 1996a,b), and the dual was not originally found constructively. (Constructive arguments were given later; see Johnson (2003); Polchinski (1998).) The $E_8 \times E_8$ theory is dual to M-theory in eleven dimensions with the extra dimension chosen to be the orbifold $\mathbf{S}/\mathbf{Z}_2$.

This leads to a new understanding of string theory in which the five distinct string theories described above are not really separate at all, but are considered as perturbation expansions of the same theory – M-theory – around different vacua, just like the perturbation expansion of a scalar theory around the extrema of its potential $V(\phi)$.

### 3.4. M-theory and the Hořava–Witten theory

Of all these models, the scenario proposed by Horava and Witten (1996b) in which the heterotic $E_8 \times E_8$ is dual to M-theory on $\mathbf{S}/\mathbf{Z}_2$ has received the most attention cosmologically. The action of $\mathbf{Z}_2$ identifies points symmetrically around $\theta = 0$ and reduces the circle to an interval bounded by two end-points. $E_8$ matter is trapped at each end-point whereas 11-dimensional supergravity (Type IIA strings) propagates in the bulk, so that when taken in total this scenario reproduces the $E_8 \times E_8$ gauge group of the heterotic string, one copy of $E_8$ for each end-point. The string coupling is still related to the size of the $\mathbf{Z}_2$. Although

the end-points are not D-branes in the sense of the previous section, they are membranes of another sort.

This model has been the basis for some novel suggestions in cosmology, which will be reviewed in the next chaper, in which our universe coincides with one of the matter-carrying end-branes, and to which the standard model gauge group is attached as some kind of GUT breaking of the $E_8$. Many (but by no means all (Lukas, Ovrut, and Waldram, 1999c)) of these models are based on the a rather arbitrary truncation of the full Hořava–Witten scenario to general relativity between the end-points and cosmological matter on the branes. Such models typically restrict attention to five dimensions, the four large dimensions of our visible universe and an extra transverse dimension corresponding to the $\mathbf{S}/\mathbf{Z}_2$ which was interpreted as the coupling of the heterotic theory. This can be justified on the basis that the characteristic size of the extra dimension must be some orders of magnitude larger than the characteristic size of the compactification manifold $K$ discussed above, which was used to wrap up the unwanted six dimensions of plain string theory. This compactification manifold is still required in M-theory compactifications, which only introduce the additional complication of the eleventh dimension.

In Horava and Witten (1996a); Witten (1996) it was shown that the low energy Newton constant arising from a compactification of the Hořava–Witten scenario to four dimensions took the form

$$G_N = \frac{\kappa^2}{16\pi^2 V r} \tag{3.4.1}$$

where $V$ is the volume of the manifold $K$, $\kappa^2$ is the eleven-dimensional gravitational coupling, and $r$ is the radius of the eleventh-dimension. Large $r$ corresponds to strong coupling, whereas when $r \to 0$, one is returning to the perturbative heterotic string. Meanwhile, the projected four-dimensional GUT coupling can be expressed as

$$\alpha_{\mathrm{GUT}} = \frac{(4\pi\kappa^2)^{2/3}}{2V}. \tag{3.4.2}$$

Remembering that $V \sim M_{\mathrm{GUT}}^{-6}$ determines the characteristic GUT scale, one can express $r$ in terms of the four-dimensional Planck scale, the four-dimensional GUT scale, and the four-dimensional GUT coupling as

$$r \sim \frac{M_{\mathrm{P}}^2}{M_{\mathrm{GUT}}^2} \alpha_{\mathrm{GUT}}^{3/2}. \tag{3.4.3}$$

The GUT coupling is known from calculations of the renormalization group, and satisfies $|\alpha_{\rm GUT}| \ll 1$. (It is probably close to some factor times $10^{-2}$.) The GUT scale must be of order $10^{16}$ GeV or so, for consistency with renormalization group calculations, and also not to disturb the success of cosmological physics such as nucleosynthesis. The Planck scale is known from precision measurements of gravity, with the result that

$$r \sim 10^{-11}\,{\rm GeV}^{-1}. \tag{3.4.4}$$

On the other hand, the characteristic scale of $K$ is $r_K \sim V^{1/6} \sim M_{\rm GUT}^{-1}\,{\rm GeV}^{-1}$, so

$$r \sim 1000 r_K, \tag{3.4.5}$$

or, $r$ is some orders of magnitude larger than the characteristic scale of $K$. This justifies treating the Hořava–Witten compactification as approximately five dimensional at low energies.

CHAPTER 4

# Cosmology

## 4.1. Introduction

In this chapter we outline the successful formulation of standard cosmology in $3+1$ spacetime dimensions and discuss some aspects of the compelling evidence which has been assembled, after rather more than thirty years of effort, to give us confidence in the idea that we can follow the evolution of the universe back in time to a hot, dense state, probably of order the Planck length, out of which the various abundant matter elements which comprise our world were extracted. For historical reasons this is often referred to as the Hot Big Bang model.

The starting point for all cosmology is the presumption that the universe appears homogeneous and isotropic on large scales. For much of the history of cosmology (as a quantitative science) it has been this supposition, rather than any of the detailed numerical predictions which flow from it and provide the framework for comparing the standard model with observation, which has been most susceptible to criticism. The general idea that the universe looks roughly the same in all directions, and from all places (but not necessarily at all times) is codified in the cosmological principle, which can naturally be viewed as the cosmological completion of the Copernican principle, that the Earth does not lie at the centre of the solar system.[1] As such, it does not follow so much from pure observational science as the *ab initio* preference for simplicity and symmetry that theoretical speculation always attracts. Over the last few years this situation has changed dramatically, with detailed mapping of significant portions of the local universe now available, out to redshifts $z \sim 1$ (Hawkins et al., 2002). When viewed in isolation small regions of this map are subject to considerable variation and exhibit much confusing noise, but, even though it is prudent to be conservative since all the evidence is not yet in, we can say with some confidence

[1] Of course, this is not quite how Copernicus enunciated his principle. The distinctions between the solar system, galaxy and universe were not quite so sharply drawn in those days.

that the universe does indeed approach homogeneity when averaged on scales somewhat larger than the size of galaxy superclusters.

There are independent estimators of the degree of isotropy and homogeneity, besides the direct but clouded route of counting galaxy distributions. Instead, one can work with the cosmic microwave background, the sea of decoupled low-energy photons which permeates the universe and survives as a relic from earlier, more inhospitable times. The existence of such a thermal background is a seemingly inevitable prediction of any model in which the universe has evolved from a hot, dense state during which equilibrium-balancing processes operated on a sufficiently small timescale to ensure that thermal equilibrium is maintained. The physics associated with the failure of thermal equilibrium is called decoupling, and will be considered later.

For a long time, comparatively little was known about the microwave background spectrum apart from from a rough estimate of its temperature and some isolated data points away from the thermal peak (for a review of early evidence, see Weinberg (1972)). A convincing demonstration of thermality had to await the data returned from the COBE mission, which found the CMB to be the most perfect thermal spectrum in nature and recorded an average temperature anisotropy

$$\left.\frac{\Delta T}{T}\right|_{\mathrm{rms}} = (1.10 \pm 0.08) \times 10^{-5}. \tag{4.1.1}$$

This has recently been improved by the WMAP experiment, as we will describe later when discussing the CMB in detail.

In models where the universe does not evolve from a hot early state, it is difficult to produce a thermal background with the required precision. Although ingenious arguments have been made regarding the synthesis of backgrounds from point sources, many difficulties remain and the existence of the CMB – with such a precise thermal spectrum – is a weighty argument against such *ad hoc* models (Albrecht, 1999; Hoyle, Burbidge, and Narlikar, 2000).

On the basis of the cosmological principle and corresponding solutions to Einstein's equations, coupled with some specification for the matter theory – it almost always suffices to deal with 'dust', or free pressureless matter, and radiation – and some straightforward aspects of thermodynamics, a great deal can be said about the subsequent evolution of the universe, the behaviour of the matter it contains, and its thermal history (Liddle and

Lyth, 2000; Peacock, 1999; Weinberg, 1972). In this chapter we describe how the various components of this structure fit together, and give a brief review of areas where the theory, or aspects of its confrontation with observation, seem less than perfect.

Having described what is more or less known, based on the available observational evidence, we then move on to a description of inflation, a conjectural extension of the standard cosmology which holds out some hope of improving on the known theoretical defects in standard cosmology: namely, the degree of flatness, homogeneity and isotropy; the dependence on initial conditions of any late-time homogeneous and isotropic phase; the generation of density fluctuations in the early universe which could grow into galaxies, clusters of galaxies, and superclusters by the present epoch; and the absence in the observable universe of exotic relics such as monopoles, cosmic strings, or other topological defects which would one would expect to generically be present in any universe which has passed through a phase transition, as our universe is conjectured to have done, from a unified gauge-theory phase at higher energy (Vilenkin and Shellard). Inflation will form a large part of the considerations of Part 2. Accessible introductions can be found in the literature (see, eg., Langlois (2004); Liddle and Lyth (1993, 2000); Lyth and Riotto (1999); Mukhanov, Feldman, and Brandenberger (1992); Peacock (1999); Riotto (2002)), but the details of some important calculations are not easy to find, and notation and nomenclature is not yet consistent across the field. For this reason, we work over the theory in some detail, and frequently attempt to give details of calculations where they are not easy to find in the literature.

Inflation is a rather generic scenario, and its predictions to some extent are independent of the precise model one chooses as a concrete realization of the inflationary concept. In view of this generality and model-independence, in order to be able to make precise statements about whole classes of models, it is important to have on hand as many exact results as possible that depend on only the most general assumptions about the conditions during inflation. As it turns out, the number of exact results which are known is unsatisfactorily small, but the most important is the no-hair theorem proved in the early 1980s by Wald (Wald, 1983). This shows that during an inflationary epoch, any structure is washed away, and the universe rapidly approaches homogeneity and isotropy within the causal horizon.

The no-hair theorem completes our preliminary survey of inflation. Before closing the present chapter, we briefly survey the observational position. The material outlined in

this rather lengthy chapter forms the basis for the treatment of brane cosmologies which is begun in the next chapter.

## 4.2. Homogeneous and isotropic cosmologies

**4.2.1. The Friedmann equation.** Any four-dimensional metric solving the Einstein equation, possibly with matter sources, and with spatial slices which are homogeneous and isotropic, must necessarily take the form (Weinberg, 1972)

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega_2^2\right), \tag{4.2.1}$$

where $\mathrm{d}\Omega_n^2$ is the metric on an $n$-sphere, and the constant $k$ characterizes the spatial curvature: $k = -1$ is an open universe, whereas $k = 0$ is flat and $k = 1$ is closed. This theorem was proved independently by Robertson and Walker, and is therefore known as the Robertson–Walker metric. However, (4.2.1) is only kinematical. The dynamics of such models were investigated by Aleksandr Friedmann, and the cosmologies arising from (4.2.1) are called Friedmann models. Given this close relationship it is quite common to conflate kinematics and dynamics, and refer to the metric (4.2.1) as the Friedman–Robertson–Walker metric or FRW metric.[2] The energy–momentum tensor describing whatever matter is coupled to gravity must also be homogeneous and isotropic, so it is guaranteed to be of the perfect fluid form

$$T^a{}_b = \mathrm{diag}(-\rho, p, p, p), \tag{4.2.2}$$

(with the present sign conventions). The Einstein equations are

$$R_{ab} - \frac{1}{2}Rg_{ab} + \Lambda g_{ab} = \kappa_4^2 T_{ab}, \tag{4.2.3}$$

[2]A brief history of relativistic cosmological models is set out in Peebles (1993). The Friedmann models (both open and closed) were first found by Friedmann, but were not widely appreciated in the West, where only the Einstein static universe and de Sitter models were commonly understood. When Slipher's results on the recession of spiral nebulae arrived, they could be interpreted in terms of the de Sitter model, where cosmological expansion occurred. On the other hand, the de Sitter model is effectively an empty universe, so it appeared to make to sense for the spiral nebulae themselves to reside outside our own galaxy. On the other hand the Einstein model was full of matter, but cosmological redshifts could not be explained within it. The resolution was supplied by Lemaître, who drew attention to the work of Friedmann, and showed that both cosmological recession and matter could be accommodated within general relativity. Some authors, for this reason, refer to (4.2.1) as the Friedmann–Lemaître–Robertson–Walker or FLRW metric.

which, when applied to (4.2.1), gives an evolution equation for $a$,

$$H^2 + \frac{k}{a^2} = \frac{\kappa_4^2}{3}\rho + \frac{\Lambda}{3} \tag{4.2.4}$$

Eq. (4.2.4) is called the Friedmann equation. Much of classical cosmology reduces just to the study of (4.2.4) with appropriate choices for $\rho$, together with some extra dynamics to take account (for example) of the evolution of perturbations, or whatever physics it is one is studying in the expanding universe. As it turns out, (4.2.4) is not really a dynamical equation at all, but the Hamiltonian constraint of general relativity, as will be described in Chapter 8. On differentiating (4.2.4), one can arrive at a dynamical equation,

$$\frac{\ddot{a}}{a} = -\frac{\kappa_4^2}{6}(\rho + 3p) + \frac{\Lambda}{3}, \tag{4.2.5}$$

which is often called the Raychaudhuri equation, by comparison with a similar equation which arises when studying the evolution of dilation and shear in general relativity (Hawking and Ellis, 1973). The Raychaudhuri equation and a conservation law for $\rho$ are together equivalent to the Friedmann equation, but since (4.2.4) is first order, most researchers work universally with it. The Raychaudhuri equation as it stands is only occasionally useful.

It is sometimes useful to rewrite the Friedmann equation in the alternative form

$$1 - \frac{k}{a^2H^2} = \frac{\kappa_4^2\rho}{3a^2H^2} + \frac{\Lambda}{3a^2H^2} = \Omega_\rho + \Omega_\Lambda = \Omega, \tag{4.2.6}$$

where the $\Omega_i$ are said to be the density parameters for the $i$th component of the cosmological fluid. The quantity $\Omega$, obtained by summing over the $\Omega_i$, is the density parameter of the universe. From (4.2.6), it is easy to see that the density parameter satisfies

$$\operatorname{sgn} k = \operatorname{sgn}(\Omega - 1). \tag{4.2.7}$$

Thus, the universe is closed, flat, or open according as $\Omega$ is greater than, equal to, or less than unity. Some authors define a curvature density $\Omega_k$ such that $\Omega_k = -k/a^2H^2$, but this definition does not seem to have much utility, and we will not employ it.

The evolution of $\Omega$ can be obtained very straightforwardly, by differentiating (4.2.6),

$$\frac{\mathrm{d}\Omega}{\mathrm{d}t} = -\frac{2k}{a^2H^2}q = 2Hq(\Omega - 1) \tag{4.2.8}$$

where $q$ is the deceleration parameter, defined by $q = -\ddot{a}a/\dot{a}^2$ (Weinberg, 1972).

**4.2.2. Thermodynamics in an expanding universe.** Having obtained the various dynamical equations which support the general relativistic discussion of cosmology, the next step is to examine the kind of matter we see around us in the local universe, and follow the physics of this material back to earlier epochs. In this section, we present a very brief account of the thermal history of the universe. The aim is not to be rigorous, but only to set a proper, observational cosmological context for the theoretical work undertaken later, and also be provide enough background to understand a survey of modern observational results. We deal only with the equilibrium theory, and we entirely ignore the effect of a cosmological constant, which (as will be seen later) is now known to be the dominant late-time constituent of the universe. At earlier epochs, however, this cosmological constant – if unevolving – would have been drowned by the sea of matter and radiation, and its effects would have been subdominant to those of ordinary types of matter, so this approximation becomes increasingly good as one follows the cosmic evolution back to the big bang. It is more difficult to justify equilibrium dynamics. In principle, one should use the full Boltzmann theory to follow the evolution, but this cannot be done analytically and one must resort to numerical codes to follow some integrations through important epochs of the universe's evolution (Peacock, 1999). Although of course this *is* what is done in practice, it is rather undesirable for a simple survey, and since the approximations turn out to be reasonably accurate, the transparency of the approximate calculations renders this simplification desirable.

More detailed treatments can be found in the literature. The classic treatment is Weinberg (1972), which describes a number of analytical treatments in considerable detail and surveys the experimental evidence which was current in the early 1970s. More recent works include Kolb and Turner (1999) and Peacock (1999), both of which address the issue of non-equilibrium dynamics and can be consulted for the more exact treatment.

The various matter components which contribute to the matter density $\rho$ on the right-hand side of Friedmann's equation (4.2.4) redshift with the expansion of the universe at rates which depend on their equation of state. Assuming that the expansion is isentropic, so that there is no net heat loss or gain as the universe expands, the first law of thermodynamics enforces the relation

$$\mathrm{d}E = -p\,\mathrm{d}V. \tag{4.2.9}$$

The energy density in a coordinate volume, in a species $i$, satisfies

$$E_i = \rho_i a^3 \tag{4.2.10}$$

whereas the coordinate volume $V$ changes according to $V = a^3$. For simple cases where the equation of state has the form $p = \omega\rho$, this gives

$$\frac{\mathrm{d}\rho_i}{\rho_i} = -3(1+\omega)\frac{\mathrm{d}a}{a}. \tag{4.2.11}$$

If $\omega$ is constant then this can be integrated at once (otherwise, one must take account of the $\rho$-dependence of $\omega$) to give

$$\rho_i = \rho_{i,0} a^{-3(1+\omega)}. \tag{4.2.12}$$

where $\rho_{i,0}$ is the present-day density, supposing that $a$ is normalised to unity at the present epoch. This allows us to follow the gross evolution of the energy density in a species $i$, and to calculate its effect on cosmological dynamics. For example, pressureless matter has $\omega = 0$ and redshifts like $\rho \propto a^{-3}$, in which case

$$\left(\frac{\dot{a}}{a}\right)^2 \propto \frac{1}{a^3}, \quad \text{so} \quad \sqrt{a}\,\mathrm{d}a \propto \mathrm{d}t. \tag{4.2.13}$$

This can be integrated at once to give $a \propto t^{2/3}$. This model is often called the Einstein–de Sitter universe. Another common equation of state is that satisfied by radiation, where $p = \rho/3$, and therefore radiation redshifts according to the law $\rho \propto a^{-4}$. The radiation-dominated scale factor behaves like $a \propto t^{1/2}$; thus, the universe grows more slowly during radiation domination, compared with matter domination.

On the other hand, it is often useful to follow the thermodynamics in more detail. The equilibrium distributions which describe the number of particles, in a multi-particle system, occupying a given single-particle state, are the Bose–Einstein distribution (appropriate for bosons) or Fermi–Dirac distributions (appropriate for fermions), parametrized in terms of the single-particle momentum $\mathbf{p}$, the mass $m$ and the energy $E = \sqrt{\mathbf{p}^2 + m^2}$. One can also include the chemical potential $\mu$, in cases where particle number is changing, and the spin degeneracy[3] $g$ which counts the number of distinct states available to a given particle of fixed $\mathbf{p}$, $m$ and $\mu$. (For example, for a photon $\gamma$, electron $e^+$, proton $p$ or neutron $n$ the

[3]We are being sloppy for notational purposes, since spin is only a meaningful concept for massive particles. In the case of massless particles such as the photon, $g$ counts helicity states rather than spin.

spin degeneracy is $g = 2$; for a scalar particle $\phi$ the spin degeneracy is $g = 1$.) The number density $n$ obeys

$$n = g \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{\exp(E-\mu)/T \pm 1} \tag{4.2.14}$$

where we have chosen units in which the Boltzmann constant $k_B = 1$, and the plus sign is taken for fermions and the minus sign for bosons. Similarly, the energy density $\rho$ is given by

$$\rho = g \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{E}{\exp(E-\mu)/T \pm 1} \tag{4.2.15}$$

and the pressure $P$ by

$$P = g \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{\mathbf{p}^2}{3E\left[\exp(E-\mu)/T \pm 1\right]}. \tag{4.2.16}$$

(We have momentarily changed notation so that the pressure $P$ is denoted with a capital $P$ to avoid confusion with the momentum $\mathbf{p}$.)

In the relativistic limit where $T \gg m, \mu$ one can expand the exponentials in terms of small powers, and integrate. For bosons one obtains

$$n = \frac{\zeta(3)}{\pi^2} g T^3, \quad \text{and} \quad \rho = \frac{\pi^2}{30} g T^4, \tag{4.2.17}$$

whereas for fermions we have

$$n = \frac{3}{4} \frac{\zeta(3)}{\pi^2} g T^3, \quad \text{and} \quad \rho = \frac{7}{4} \frac{\pi^2}{30} g T^4. \tag{4.2.18}$$

In both cases the pressure satisfies $P = \rho/3$, which is the equation of state for a relativistic ideal gas. The non-relativistic limit occurs where the temperature $T$ is much less than the rest mass $m$. Since

$$E = (\mathbf{p}^2 + m^2)^{1/2} \approx m + \frac{\mathbf{p}^2}{2m}, \tag{4.2.19}$$

we have

$$n = g \exp\left(-\frac{m-\mu}{T}\right) \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \exp\left(-\frac{\mathbf{p}^2}{2m}\right) = g \left(\frac{mT}{2\pi}\right)^{3/2} \exp\left(-\frac{m-\mu}{T}\right). \tag{4.2.20}$$

The density $\rho$ is given by $\rho = mn$, and the pressure $P$ by $P = nT \ll \rho$. This formula shows that, for $\mu \approx 0$, non-relativistic massive particles are exponentially suppressed whenever $T < m$.

We define the temperature $T$ of the universe to be the photon temperature, $T \equiv T_\gamma$ and enumerate the various other matter species in the universe by the label $i$, defining appropriate temperatures $T_i$, pressures $P_i$ and densities $\rho_i$. Where microphysical processes serve to bind any of these constituents in thermal equilibrium with the photons, the temperature

$T_i$ will equal the photon temperature $T$. These microphysical process typically have activation energies which depend on the temperature, and as the universe cools, one by one the interactions begin to peter out, and the various other species $i$ present in the cosmic fluid fall out of equilibrium with the photons. During any period where the expansion of the universe is isentropic, the photon temperature falls only owing to gravitational interactions in which the cosmic expansion redshifts individual photon energies, but at decoupling thresholds or where strongly coupled particle interactions take place the expansion may not be thermodynamically reversible and appreciable entropy may be deposited in the thermal bath, leading to an effective increase in $T$. This process is called reheating. After falling out of equilibrium with the photons, each species $i$ continues with the temperature $T_i$ it had at decoupling, cooled by the expansion of the universe, but takes no further part in interactions with the ambient thermal bath, and therefore is unaffected by reheating, so in general $T_i \neq T$ once further entropy production has occurred.

The total energy density in relativistic species can be written as an effective Stefan–Boltzmann law

$$\rho_r = \frac{\pi^2}{30}\mathcal{N}(T)T^4, \tag{4.2.21}$$

where $\mathcal{N}(T)$ is a sum over the effective relativistic degrees of freedom, consisting of two components:

- For relativistic species $i$ in thermal equilibrium, $T_i = T$, so the sum over degrees of freedom in the thermal bath $\mathcal{N}_*(T)$ is

$$\mathcal{N}_*(T) = \sum_{\text{bosons}} g_i + \frac{7}{8}\sum_{\text{fermions}} g_i, \tag{4.2.22}$$

  since fermions contribute to the energy density like 7/8 of a particle.

- For relativistic species $i$ *not* in thermal equilibrium, the temperature $T_i$ may be different from $T$, so the sum of decoupled degrees of freedom is

$$\mathcal{N}_D(T) = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8}\sum_{\text{fermions}} g_i \left(\frac{T_i}{T}\right)^4. \tag{4.2.23}$$

The total number of relativistic degrees of freedom is $\mathcal{N}(T) = \mathcal{N}_*(T) + \mathcal{N}_D(T)$, which is constant away from decoupling thresholds.

If radiation domination is a good approximation, we can compare the energy density in relativistic species with the energy density in the radiation dominated Einstein–de Sitter

model (4.2.13) to find

$$\frac{\pi^2}{30}\mathcal{N}(T)T^4 = \frac{3}{32\pi G t^2}, \tag{4.2.24}$$

and rearrange for $t$

$$t = \left(\frac{90}{32\pi^3 G\mathcal{N}(T)}\right)^{1/2} T^{-2} \simeq 0.30\mathcal{N}(T)^{-1/2}\frac{M_{\rm P}}{T^2} \tag{4.2.25}$$

In terms of the radiation temperature, the Hubble parameter satisfies

$$H \simeq 1.66\mathcal{N}(T)^{1/2}\frac{T^2}{M_{\rm P}}. \tag{4.2.26}$$

These considerations serve to identify the evolution of $T$ in the large, or whenever $\mathcal{N}(T)$ is not changing. Since one can show, in the full general relativistic formulation, that the expansion of the universe is overwhelmingly isentropic,[4] we can probe the variation of $T$ as the universe expands and various species $i$ decouple using entropy arguments. To obtain this, we employ the integrated first law, which is known as Euler's equation,

$$TS = U + PV - \sum_i \mu_i N_i. \tag{4.2.27}$$

The chemical potential $\mu_i$ satisfies $\mu_i \approx 0$, so the Euler equation can be used to compute the entropy density $s_i$,

$$s_i = \frac{S_i}{V} = \frac{1}{T_i}(\rho_i + P_i) = \frac{4}{3}\frac{\rho_i}{T_i}, \tag{4.2.28}$$

where we have used the relativistic equation of state $P_i = \rho_i/3$ and expressed the internal energy in terms of the relativistic energy density, giving $U = V\rho_i$. By analogy with the total energy density in relativistic species, one can define a total entropy density $s$ by

$$s = \sum_i s_i = \frac{2\pi^2}{45}\mathcal{N}_S(T)T^3 \tag{4.2.29}$$

where $\mathcal{N}_S(T)$ has thermal and non-thermal contributions, as for the energy density,

$$\mathcal{N}_S(T) = \left[\sum_{\rm bosons} g_i + \frac{7}{8}\sum_{\rm fermions} g_i\right] + \left[\sum_{\rm bosons} g_i\left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{\rm fermions} g_i\left(\frac{T_i}{T}\right)^3\right]. \tag{4.2.30}$$

If the total entropy $S \propto a^3\mathcal{N}_S(T)T^3$ is conserved, then

$$T \propto \mathcal{N}_S(T)^{-1/3}a^{-1}. \tag{4.2.31}$$

[4]The expansion might not be exactly isentropic, because of irreversible processes occurring on microscopic scales. However, this small scale entropy change is enormously dwarfed by the bulk entropy of the cosmic microwave background, for which isentropy is a good approximation. It is therefore quite justifiable to ignore small scale irreversibilities when considering the gross evolution of the universe.

Notice that this needs treating with caution near decoupling thresholds, where $\mathcal{N}_S(T)$ may change.

**4.2.3. Decoupling and freeze-out.** A species $i$ is maintained in thermal equilibrium by interactions with other particles in the thermal bath. The rate $\Gamma_i$ for these interactions is usually governed by a formula of the form

$$\Gamma_i = n_j \sigma \bar{v} \tag{4.2.32}$$

where $n_j$ is the number density of the species $j$ interacting with $i$, and $\sigma$ is the cross-section for the reaction. The parameter $\bar{v}$ determining the overall scale of the interaction is determined by a suitably averaged particle velocity.

- Massive gauge bosons. These have mass $m_i > T$ and interact weakly with gauge coupling $g = \sqrt{4\pi\alpha}$ and have propagator proportional to $G_i = \alpha/m_i^2$. On dimensional grounds, we expect the cross-section to have the form

$$\sigma \sim G_i^2 T^2 \tag{4.2.33}$$

  Using $n_{e^+} = 0.2T^3$, appropriate for the number density of electrons as the interacting species, we find

$$\Gamma_{e^+} \sim G_A^2 T^5 \tag{4.2.34}$$

- Massless gauge bosons. The cross-section here has the form $\sigma \sim \alpha^2/T$, so assuming the electron number density again (this is the particle the gauge bosons are interacting with, so there is no inconsistency in using it twice), we have

$$\Gamma_{e^+} \sim \alpha^2 T \tag{4.2.35}$$

  An example of a massless gauge boson is the photon $\gamma$.

- Non-relativistic species. These are exponentially suppressed when in thermal equilibrium, so $\Gamma_i$ decreases rapidly.

Decoupling or 'freeze-out' of a particular interaction occurs whenever the interaction rate $\Gamma$ for that process falls below the Hubble expansion rate. When this happens, the expected time $1/\Gamma$ between interactions is greater than the expansion time $1/H$ of the universe, so we do not, on average, expect particles to interact again during the lifetime of the universe. Decoupling is a sharp transition: in the régime $\Gamma > H$, couplings between particle keep the species $i$ tightly bound in thermal equilibrium, but when $\Gamma < H$, these

very same particles are not expected to interact again during the lifetime of the universe. The decoupling temperature for species $i$ is defined as the temperature when $\Gamma_i = H$.

**Example.** The canonical example of decoupling is provided by neutrinos. For $T \gg$ MeV, weak interactions maintain the neutrinos in thermal equilibrium, for example

$$\begin{aligned} e^- + \nu_e &\longleftrightarrow e^- + \nu_e \\ e^- + e+ &\longleftrightarrow \nu_e + \bar{\nu}_e \end{aligned} \tag{4.2.36}$$

and so on. The Fermi cross-section for these interactions is $\sigma_W = G_F^2 T^2$ (as we saw in the case of massive gauge bosons described above), where $G_F$ is the Fermi coupling

$$G_F = 1.17 \times 10^{-5}\ \mathrm{GeV}^{-2} \simeq 10^{-5} m_p^{-2}. \tag{4.2.37}$$

Here $m_p$ is the proton mass, not the Planck scale. The important species for keeping neutrinos in equilibrium are other neutrinos and electrons, and their antiparticles, so the number density of particles with which a given neutrino can interact is given by

$$n = n_e + n_\nu = 0.2T^3 + 0.1T^3 = 0.3T^3 \tag{4.2.38}$$

taking account of the two spin states of the electron, and the single spin state of a neutrino. Thus, the interaction rate approximates

$$\Gamma_{\nu_e} \simeq 0.3 G_F^2 T^5 \tag{4.2.39}$$

To find the decoupling temperature, one compares the neutrino interaction rate (4.2.39) with the expansion rate (4.2.26), for which it is necessary to know the number of degrees of freedom $\mathcal{N}(T)$ of relativistic species. The important contributions to $\mathcal{N}(T)$ come from photons, which contribute 2 helicity degrees of freedom; electrons and positrons, which contribute two spin degrees of freedom each; and the six types of neutrino, which contribute one spin degree of freedom each. Remembering that fermions contribute like 7/8 of a particle, we have

$$\mathcal{N}(T) = 2[\gamma] + \frac{7}{8}\left(4[e^\pm] + 6[\nu, \bar{\nu}]\right) = 10.75, \tag{4.2.40}$$

so we can approximate

$$\frac{\Gamma_{\nu_e}}{H} = \frac{10^{-5}}{1.66\sqrt{10.75}} \frac{T^3}{m_p^2 M_{\mathrm{P}}} \simeq \left(\frac{T}{2\,\mathrm{MeV}}\right)^3. \tag{4.2.41}$$

Thus electron-type neutrinos decouple at a physical temperature $T_{\nu_e} \simeq 2\,\mathrm{MeV}$. A full Boltzmann analysis gives $T_{\nu_e} \approx 1\,\mathrm{MeV}$, so this is not a bad estimate.[5]

[5] see the discussion of the Boltzmann equation and decoupling in Kolb and Turner (1999); see also Cowsik and McClelland (1972); Weinberg (1975); Zel'dovich (1966). In fact, one can be rather more severe. It is possible to derive limits on the *mass* of the neutrino species from cosmological arguments, and in fact this is what was done in the Cowsik–McClelland paper; the bound for light neutrinos is known as the Cowsik–McClelland bound. For massive neutrinos it is called the Lee–Weinberg bound.

**4.2.4. Recombination.** Once $T$ has dropped sufficiently, protons and electrons can combine to form hydrogen. This is the process

$$p + e^- \longleftrightarrow H + \gamma \tag{4.2.42}$$

which is often called recombination, despite the fact that $p$ and $e^-$ have in fact never been in association before. At high temperatures, whenever a chance association occurs, it is instantaneously disrupted by impact from a background photon with energy greater than the ionization energy of hydrogen. Since this process is occurring in equilibrium, the chemical potentials of electrons and protons must be related via[6]

$$\mu_H = \mu_p + \mu_{e^-}. \tag{4.2.44}$$

In addition, charge neutrality requires that $n_e = n_p$. To find the ratio of hydrogen to protons and electrons, one divides the non-relativistic equilibrium distributions for $n_H$, $n_p$ and $n_e$,

$$\frac{n_H}{n_e n_p} = \frac{g_H}{g_p g_e}\left(\frac{2\pi m_H}{m_e m_p T}\right)^{3/2} \exp\left(-\frac{m_H - m_p - m_e}{T}\right), \tag{4.2.45}$$

where we have eliminated $\mu$s using the equilibrium relation. The mass difference $m_H - m_e - m_p$ is just the ionization energy $I = 13.6\,\mathrm{eV}$ for hydrogen, and the ratio of spin degeneracies is unity, since $g_H = 4$. Therefore,

$$\frac{n_H}{n_p n_e} \approx \left(\frac{2\pi}{m_e T}\right)^{3/2} \exp\left(\frac{I}{T}\right), \tag{4.2.46}$$

where we have approximated $m_H \approx m_p$. Define a fractional ionization $X_e$ as

$$X_e = \frac{n_e}{n_B} = \frac{n_p}{n_B} \tag{4.2.47}$$

(where the last equality follows from charge neutrality), with the baryon number $n_B$ defined by

$$n_B = n_p + n_H = \eta n_\gamma. \tag{4.2.48}$$

The ratio $\eta$ which describes the relative populations of baryons and photons is known as the baryon to photon ratio. Although one might expect the number density of neutrons $n_n$ to appear in the baryon number, it is a good approximation to ignore it, since almost

[6]The photon chemical potential is zero, $\mu_\gamma = 0$. This follows from processes such as Bremsstrahlung,

$$e^- + \gamma \longrightarrow e^- + \gamma + \gamma \tag{4.2.43}$$

all neutrons are captured into $He^4$ nuclei before this time, although we have ignored such nuclei in the calculation. Since the photon number density $n_\gamma$ is known, this provides an estimate of $n_B$,

$$n_B = \eta \frac{2\zeta(3)}{\pi^2} T^3. \tag{4.2.49}$$

Substituting in (4.2.46) gives Saha's equation, which describes the equilibrium balance between abundances of protons, electrons and elemental hydrogen:

$$\frac{1 - X_e}{X_e} = \frac{2\zeta(3)}{\pi^2} \eta \left( \frac{2\pi T}{m_e} \right)^{3/2} \exp\left( \frac{I}{T} \right). \tag{4.2.50}$$

Like any equilibrium process, recombination never goes to completion. Instead, varying abundances of H, $p$ and $e^+$ are dynamically adjusted to suit the equilibrium conditions; at high energy, $p$ and $e^+$ are favoured, whereas at low energy, elemental $H$ is profitable. Therefore it is reasonable to adopt a condition where recombination is supposed to be reasonably complete, such as when the fractional ionization $X_e$ falls below 0.1. The temperature of recombination $T_{\rm rec}$ was then

$$T_{\rm rec} = T_0(1 + z_{\rm rec}) \simeq 3600\,{\rm K} = 0.31\,{\rm eV}. \tag{4.2.51}$$

The photon-to-baryon ratio $\eta$ can be measured from present-day experiments[7] and satisfies $\eta \approx 10^{-9}$. $T_{\rm rec}$ is much lower than the naïve expectation that the recombination temperature should be close to the ionization temperature of hydrogen. However, this is not really any great surprise since processes such as nuclear fusion in stars, in which the long tail of the Maxwell–Boltzmann distribution can support high energy particles even at relatively low ambient temperatures, are familiar and important. The same physics applies here: recombination is more or less controlled by the Maxwell–Bolzmann tail.

**4.2.5. Primordial nucleosynthesis.** As the temperature falls still further, protons and neutrons can combine to form the nuclei of light elements, a process known as nucleosynthesis. At one time it was thought that all elements might be formed in this way, but it is now understood that only relatively light elements, such as H, He, and Li can be formed from fusion of $n$ and $p$ after the big bang, whereas heavier elements like C, necessary for life on earth, must be formed in stars.

[7] In the past, this was often done via constraints from cosmological nucleosynthesis, but is today best achieved from cosmic microwave background experiments (Spergel et al., 2003).

The neutron–proton mass difference will suppress neutrons with respect to the lighter protons before decoupling, when both species are in the non-relativistic régime and subject to the Maxwell–Boltzmann suppression described by (4.2.20). The important reactions maintaining equilibrium are

$$\begin{aligned} n + \nu &\longleftrightarrow p + e^- \\ n + e^+ &\longleftrightarrow p + \bar{\nu} \end{aligned} \tag{4.2.52}$$

and the neutron decay reaction ($\beta$ decay)

$$n \longrightarrow p + \bar{\nu} + e^-. \tag{4.2.53}$$

Assuming chemical equilibrium, so that the potentials $\mu_i$ are approximately zero, and charge neutrality, so that $n_e \approx n_p$,

$$\frac{n_n}{n_p} = \frac{X_n}{X_p} = \exp\left(-\frac{Q}{T}\right) \tag{4.2.54}$$

where $Q$ is the mass difference $m_n - m_p = 1.29\,\mathrm{MeV}$, and the number density fraction $X_i$ is

$$X_i = \frac{n_i}{n_B}, \tag{4.2.55}$$

the number density ratio of species $i$ to the baryons. This is usually a convenient parametrization for number densities. To actually fix the neutron to proton ratio, it is necessary to know the temperature at which neutrons and protons decouple, after which the ratio $n_n/n_p$ is frozen in. The equilibrium processes (4.2.52) all occur on weak timescales for which $\Gamma_W \simeq G_F^2 T^5$, so repeating the thermal argument we applied to neutrinos above, one has freeze-out for

$$\frac{\Gamma_W}{H} \simeq \left(\frac{T}{0.8\,\mathrm{MeV}}\right)^3. \tag{4.2.56}$$

Thus, protons and neutrons decouple at temperatures of order $T \approx 0.8\,\mathrm{MeV}$. Using this figure in the equilibrium equation shows that, at decoupling, the proton-to-neutron ratio is approximately

$$\left.\frac{X_n}{X_p}\right|_{\text{decoupling}} \simeq \frac{1}{6}. \tag{4.2.57}$$

Meanwhile, neutron decay continues up until the time of nucleosynthesis, beginning with deuterium formation at about $T = 0.1\,\mathrm{MeV}$. When freeze-out occurs, this is about 100s away. Neutron decay can be approximated by a stochastic process,

$$X_n = X_{n,\text{decoupling}} \exp\left(-\frac{t}{\tau_n}\right) \tag{4.2.58}$$

where the characteristic decay time $\tau_n$ is of the order of 10 mins. This conversion of neutrons to protons means that at the beginning of nucleosynthesis, the neutron-to-proton ratio will be about $X_n/X_p \simeq 1/7$.

Because of the high plasma temperature, three-body interactions are almost negligible and nucleosynthesis is instead constrained to proceed via the intermediate deuterium-synthesis reaction

$$p + n \longleftrightarrow d + \gamma. \tag{4.2.59}$$

However, the binding energy for deuterium is low, only roughly 2.2 MeV, so substantial deuterium production only occurs well below temperatures where stable He can form. (The binding energy for He is about 28.3 MeV). This is called the deuterium bottleneck. The ratio of the fractional density of deuterium to neutrons and protons in isolation can be written, assuming equilibrium dynamics,

$$\frac{X_d}{X_p X_n} = 16.3 \left(\frac{T}{m_p}\right)^{3/2} \eta \exp\left(\frac{B_d}{T}\right) \tag{4.2.60}$$

in analogy with the Saha equation for hydrogen abundance at recombination. At $T \simeq 0.1$ MeV, large-scale deuterium production can begin, which then cascades almost immediately into $He^4$. Almost all neutrons are captured in this process.

The abundance by mass of $He^4$, written $Y_p$, can be calculated *approximately* via

$$Y_p = \frac{m_{He^4}}{m_H} X_{He^4} \tag{4.2.61}$$

if we assume that all neutrons are captured into $He^4$ and any remaining protons recombine to form H. The quantity $X_{He^4}$ is then the relative numerical abundance of Helium with respect to hydrogen. Since it takes two neutrons to construct a $He^4$ nucleus, we have $X_{He^4} = X_n/2$, and

$$Y_p = 4 \times \frac{X_n}{2} = 2X_n = \frac{2 \times \frac{1}{7}}{1 + \frac{1}{7}} = 0.25. \tag{4.2.62}$$

This is in remarkably good agreement with the observed value $Y_p = 0.246 \pm 0.014$.

## 4.3. Large scale structure formation

Having understood the gross evolution of the universe, it is now necessary to include the effect of small perturbations. This is important for two reasons. Firstly, the homogeneous and isotropic cosmologies described in the preceding section are entirely uniform and can never describe the universe we see around us, which contains pockets of high and low

matter density. We live in an appreciable condensation, but elsewhere there are voids where space is almost empty of matter, and clusters where the matter density is much greater than that we see around us today. The second, and more important, reason is that it is only through the study of perturbations that we can make contact with observation. Many of the observations available to cosmologists rely on the study of galaxy clustering, or the distribution of gas, or relative anisotropies in number counts or densities of objects on the sky: knowing the background level is not enough.[8] Instead, one must study how objects cluster, group and evolve as a function of redshift. For these reasons, and, most importantly, to study galaxy clustering via the 2dF or Sloan surveys and to study CMB anisotropies described by COBE and WMAP, one must study perturbation theory.

**4.3.1. The density perturbation.** The density perturbation is defined by

$$\delta(\mathbf{r},t) = \frac{\rho(\mathbf{r},t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{4.3.1}$$

where $\bar{\rho}(t)$ is the average density for a given homogeneous and isotropic FRW model which serves as the cosmological background, and $\rho(\mathbf{r},t)$ is the full density, including some perturbation, which may be spatially dependent and need not be homogeneous or isotropic. For convenience it is easiest to work in flat model where distractions owing to the geometry of the universe are minimal, and in fact this model provides the closest approximation to reality. One expands $\delta$ in comoving Fourier modes,

$$\delta(\mathbf{r},t) = \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \delta(\mathbf{k},t) \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}, \tag{4.3.2}$$

where $\delta(\mathbf{k},t)$ is given by

$$\delta(\mathbf{k},t) = \int \frac{\mathrm{d}^3r}{(2\pi)^{3/2}} \delta(\mathbf{r},t) \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}}. \tag{4.3.3}$$

Here, $\mathbf{k}$ is the comoving wavevector with physical wavelength $\lambda$, which scales with $a$,

$$\lambda(t) = \frac{2\pi a(t)}{k} \quad \text{where} \quad k = |\mathbf{k}|. \tag{4.3.4}$$

[8] Of course, it is not always the case that predictions in cosmology are restricted to the results of perturbation theory. For example, one may point to the successful prediction of the primordial $^4$He abundance or the general machinery of thermodynamics in the expanding universe, which are certainly non-perturbative and in excellent agreement with observation. These $O(1)$ results were mostly worked out in the early days and are summarized in quite some detail in Weinberg's influential monograph (Weinberg, 1972). However, once gravitational processes on small scales are involved, our ability to calculate exact results evaporates and one is forced to rely on perturbative approaches.

The power spectrum $\mathcal{P}(k)$ is defined by[9]

$$\mathcal{P}(k) = \langle |\delta(\mathbf{k}, t)|^2 \rangle, \tag{4.3.8}$$

which is the Fourier transform of the galaxy two-point correlation function $\xi(\mathbf{r})$. This is sometimes known as the Wiener–Khinchin theorem, even though it is in fact trivially a consequence of the convolution theorem for Fourier transforms. Many models of structure formation predict a scaling behaviour for $\mathcal{P}(k)$ like

$$\mathcal{P}(k) = Ak^n \tag{4.3.9}$$

where $n$ is a characteristic number known as the matter spectral index. The aim for models of large scale structure is to predict the power spectrum *today*, written $\mathcal{P}_0(k)$, for a given cosmology $\Omega_{\mathrm{B}}$, $H_0$, $\Omega_{\mathrm{CDM}}$, $\Omega_\Lambda$ (and so on), given the primordial form of $\mathcal{P}(k)$, written $\mathcal{P}_\infty(k)$, which must be supplied separately. The calculation of the primordial power spectrum is the province of models of the early universe. In this chapter, we only consider inflation, and derive the primordial power spectrum resulting from an epoch of scalar-field driven early universe inflation in (4.6.82); but in later chapters we will consider the calculation of a primordial $\mathcal{P}_\infty(k)$ in other possible theories of the early universe.

---

[9]This equation is not precisely correct as written, unless one makes the expansion in finite volume, in other words replacing Fourier integrals by Fourier series. What is at issue is the formalization of the intuitive idea that the power in the perturbation $\delta$ should be related to its square $\delta^2$,

$$\delta(\mathbf{r})^2 = \int \frac{\mathrm{d}^3k\,\mathrm{d}^3k'}{(2\pi)^3}\delta(\mathbf{k})\delta(\mathbf{k}')\mathrm{e}^{\mathrm{i}\mathbf{r}\cdot(\mathbf{k}+\mathbf{k}')}. \tag{4.3.5}$$

When averaged over all space, $\int \mathrm{d}^3r$, this says, up to normalization,

$$\langle \delta(\mathbf{r})^2 \rangle_{\mathrm{r}} = \int \mathrm{d}^3k\ \delta(\mathbf{k})\delta(-\mathbf{k}) = \int \mathrm{d}^3k\ |\delta(\mathbf{k})|^2, \tag{4.3.6}$$

since $\delta(\mathbf{r})$ is real and its Fourier transform therefore obeys the reality condition $\delta(-\mathbf{k}) = \delta(\mathbf{k})^\dagger$. The power spectrum is, as usual, the power per $\mathbf{k}$-space interval. This seems to justify the form of the power spectrum which has been given, but it is necessary to be careful about the averaging involved. It is quite usual to suppose that the density perturbation is the realization of an ergodic process, in the sense that one can legitimately trade ensemble averages for spatial averages (when averaging over causally disconnected cells this may indeed be literally true without the need to invoke any ergodic theorem). Since the Fourier modes $\delta(\mathbf{k})$ are uncorrelated this shows that we should in fact more carefully write

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle_{\mathrm{ensemble}} = P(k)\delta_{\mathrm{D}}(\mathbf{k} - \mathbf{k}'), \tag{4.3.7}$$

again up to normalization.

**4.3.2. Perturbation theory.** Many equations in cosmology take on considerably simpler forms when written in terms of a conformal time $\tau$, which is defined by $\mathrm{d}t = a\,\mathrm{d}\tau$. There is nothing mysterious about this arrangement; when written in terms of the conformal time, the FRW metric is conformally flat, so many equations are related to their flat counterparts via only conformal transformations. To do perturbation theory we include a small disturbance away from flatness, which is most easily written as a fluctuation in the flat metric, so one obtains

$$\mathrm{d}s^2 = a^2(\tau)(\eta_{ab} + h_{ab})\mathrm{d}x^a\,\mathrm{d}x^b. \tag{4.3.10}$$

As was outlined in Appendix B, general relativity is a gauge theory, with gauge group $GL_4$, or $SO(3,1)$ in the vielbein formalism. Therefore $h_{ab}$ is a $GL_4$ matrix, and contains gauge degrees of freedom that relate numerically distinct but physically equivalent $h_{ab}$ via $GL_4$ transformations, and to remove this redundancy and actually calculate we shall have to impose some extra conditions on $h_{ab}$, or fix the gauge. A convenient choice (but somewhat old-fashioned) is the synchronous gauge in which $t$ remains the proper time measured by a comoving observer. Therefore,

$$h_{00} = 0, \quad \text{and} \quad h_{0i} = 0 \quad \text{(synchronous gauge)}. \tag{4.3.11}$$

In modern work, it is rather more fashionable (and convenient) to work either with gauge-invariant observables, or sometimes the generalized Newtonian gauge in which the metric takes the form, for a flat universe,

$$\mathrm{d}s^2 = a(\tau)^2\left[-(1+2\Phi)\mathrm{d}\tau^2 + (1-2\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right], \tag{4.3.12}$$

where $\Phi$ is essentially the Newtonian gravitational potential in the limit $c^{-1} \downarrow 0$, and $\Phi = \Psi$ where there is no anisotropic stress. Either of these modern choices brings considerable advantages. In principle, the gauge-invariant method is best, not least because it eliminates at a stroke the complicated issues of interpretation which occur in gauge-fixed approaches, where one must be certain that the effects one seems to see are real and not gauge artefacts. On the other hand, the gauge-invariant approach can be calculationally intensive, because one must include all perturbation modes of the metric, including off-diagonal and tensor pieces. The conformal Newtonian gauge is mathematically simple and leads to a much reduced calculational complexity. Indeed, one can profitably combine these methods

(Liddle and Lyth, 2000), calculating quantities in the conformal Newtonian gauge which can then be translated to gauge-invariant observables.

We will describe the gauge-invariant formalism when discussing the generation of density fluctuations from inflation, so in this section we choose a gauge-fixed formalism to reduce the amount of calculation involved, and also to provide some contrast and a basis for comparison. The synchronous gauge, although old-fashioned, is rather transparent in this context because the time coordinate remains unperturbed, and, of course, the answers we obtain will agree with the same calculation done in any other gauge, or in a gauge-free manner.

There is another choice, the fluid-flow formalism which is outlined in its most mature form by Liddle and Lyth (1993, 2000). This method is appealing to many cosmologists, since it works with locally defined perturbations in the physical constituents of the universe, $\rho$ and $p$, and a locally defined Hubble rate, and the perturbation equations are essentially only the equations of relativistic hydrodynamics. Exactly which method is preferred is something of a matter of personal taste. In this thesis not much use is made of the fluid-flow formalism, and we prefer to work in terms of the perturbed metric tensor, a rather more 'field-theoretic' technique.

For a multicomponent fluid, where the various component fluids are labelled $i$, we have an energy–momentum tensor of the form

$$T^{ab} = \sum_i \left[ (\rho_i + P_i) u_i^a u_i^b - P_i g^{ab} \right] \tag{4.3.13}$$

To rewrite this in terms of the perturbation, we express the density as

$$\rho_i(\mathbf{r}, t) = \bar{\rho}_i(t) \left[ 1 + \delta_i(\mathbf{r}, t) \right], \tag{4.3.14}$$

where $\bar{\rho}$ denotes the average background density, and similarly express the pressure as

$$P_i = \omega_i \rho_i = \bar{\rho}_i \omega_i \left( 1 + \delta_i \right) \tag{4.3.15}$$

where $\omega_i$ parametrizes the equation of state for the $i$th fluid, $P_i = \omega_i \rho_i$, and is assumed to be a constant. (For example, $\omega_i = 1/3$ for radiation and zero for matter.) The four-velocities $u^a$ satisfy

$$u_i^a = \bar{u}_i^a + \delta u_i^a = \frac{1}{a}(1, 0, 0, 0) + \frac{1}{a}(0, \mathbf{v}_i), \tag{4.3.16}$$

with $|\mathbf{v}_i| \ll 1$. To first order, the energy–momentum tensor is

$$\begin{aligned} T_{00} &= \frac{1}{a^2}\sum_i \bar{\rho}_i(1+\delta_i) = \frac{3M_{\rm P}^2}{8\pi}\left(\frac{a'}{a}\right)^2 \sum_i \Omega_i(1+\delta_i) \\ T_{\alpha\beta} &= \frac{3M_{\rm P}^2}{8\pi}\left(\frac{a'}{a}\right)^2 \sum_i \omega_i\Omega_i\left[(1+\delta_i)\delta_{\alpha\beta} + h_{\alpha\beta}\right] \\ T_{0\beta} &= \frac{3M_{\rm P}^2}{8\pi}\left(\frac{a'}{a}\right)^2 \sum_i (1+\omega_i)\Omega_i v_i^\beta \end{aligned} \tag{4.3.17}$$

where $\alpha$, $\beta$ refer to coordinate in the three-geometry, and indices are raised and lowered with the Minkowski metric. In particular, therefore, there is no distinction between $v^\alpha$ and $v_\alpha$, so we place the index conveniently. Throughout, the background density $\bar{\rho}_i$ has been replaced by the density parameter $\Omega_i$, which is related to $\bar{\rho}_i$ via

$$\bar{\rho}_i = \frac{3M_{\rm P}^2}{8\pi}\left(\frac{a'}{a}\right)^2 \Omega_i. \tag{4.3.18}$$

Conservation of energy–momentum applied to these expressions gives two evolution equations,

$$\begin{aligned} \delta_i' + (1+\omega_i)\mathrm{i}\mathbf{k}\cdot\mathbf{v}_i - \frac{1}{2}(1+\omega_i)h' &= 0 \\ \mathbf{v}_i' + (1-3\omega_i)\frac{a'}{a}\mathbf{v}_i + \frac{\omega_i}{1+\omega_i}\mathrm{i}\mathbf{k}\delta_i &= 0. \end{aligned} \tag{4.3.19}$$

where we have translated to Fourier space and set $h = \mathrm{Tr}\, h$.

**Example.** Cold dark matter is the simplest case, corresponding to non-relativistic pressureless matter in free expansion. There is no pressure or velocity perturbation. Therefore,

$$\delta_c' = \frac{h'}{2} \quad \text{or, with appropriate initial conditions} \quad \delta_c = \frac{h}{2} \tag{4.3.20}$$

This is just a density effect to due changes in proper volume, since $\sqrt{\det g_{\alpha\beta}} \simeq a^3(1-h/2)$.

**Example.** Another important example is radiation, which is needed for the discussion of CMB anisotropies. Here $\omega_r = 1/3$, so the evolution equations for the density and velocity perturbations become

$$\begin{aligned} \delta_r' + \frac{4}{3}\left(\mathrm{i}\mathbf{k}\cdot\mathbf{v}_r - \frac{h'}{2}\right) &= 0 \\ \mathbf{v}_r' + \frac{1}{4}\mathrm{i}\mathbf{k}\delta_r &= 0. \end{aligned} \tag{4.3.21}$$

These can be combined into a single evolution equation for $\delta_r$, in which $\mathbf{v}_r$ has been eliminated,

$$\delta_r'' + \underbrace{\frac{1}{3}k^2\delta_r}_{\text{pressure}} - \frac{2}{3}h'' = 0. \tag{4.3.22}$$

These conservation laws allow us to express $\delta$ in terms of $h$ and other geometrical quantities, but there remains the problem of solving for $h_{ij}$ itself. For this purpose, there are also evolution equations arising from the linearised Einstein field equations. The most important of these is a trace equation,

$$h'' + \frac{a'}{a}h' - 3\left(\frac{a'}{a}\right)^2 \sum_i (1+3\omega_i)\Omega_i\delta_i = 0. \tag{4.3.23}$$

**Example.** Consider cold dark matter coupled to radiation. Bearing in mind that $\delta_c = h/2$ (with appropriate initial conditions; see Eq. (4.3.20) and the discussion of adiabaticity below), the trace equation (4.3.22) is

$$\delta_c'' + \frac{a'}{a}\delta_c' - \frac{3}{2}\left(\frac{a'}{a}\right)^2 (\Omega_c\delta_c + 2\Omega_r\delta_r) = 0 \tag{4.3.24}$$

Meanwhile, the radiation evolution equation is just (4.3.22)

$$\delta_r'' + \frac{1}{3}k^2\delta_r - \frac{4}{3}\delta_c'' = 0. \tag{4.3.25}$$

There are two important cases.

- Superhorizon evolution. This is the situation where the wavelength of a particular mode is larger than the horizon scale, which is $k\tau < 2\pi$. Pressure effects can be ignored ($k^2\delta_r/3 \ll \delta_r''$) because the sound speed will never be large enough to cross the perturbation, compared with the timescale for gravitational instability.
- Subhorizon evolution. This occurs when the wavelength is smaller than the horizon scale, $k\tau > 2\pi$. Pressure is important in this case.

The evolution of perturbations depends on the cosmological background.

- Matter era. Here $\Omega_{\rm CDM} \approx 1$, $\Lambda$ domination has yet to occur, and radiation is negligible. Therefore,

$$\delta_c'' + \frac{a'}{a}\delta_c' - \frac{3}{2}\left(\frac{a'}{a}\right)^2 \delta_c = 0. \tag{4.3.26}$$

  In the matter era, $a \propto t^{2/3}$, so $a'/a = 2/\tau$, leaving

$$\delta_c'' + \frac{2}{\tau}\delta_c' - \frac{6}{\tau^2}\delta_c = 0. \tag{4.3.27}$$

  There is a standard technique for solving equations of this sort, which will also be useful in the case of radiation domination below. Firstly, note that this is an equidimensional equation, so the solution is of the form $\delta_c \propto \tau^\lambda$ for some $\lambda$ which must obey the indicial equation,

$$\lambda(\lambda-1) + 2\lambda - 6 = 0 \quad \Leftrightarrow \quad \lambda^2 + \lambda - 6 = 0. \tag{4.3.28}$$

This is found by direct substition in (4.3.26). The roots lie at $\lambda = 2, -3$, so in conformal time the solution has the form

$$\delta_c = \tilde{A}(\mathbf{k}) \left( \frac{\tau}{\tau_\infty} \right)^2 + \tilde{B}(\mathbf{k}) \left( \frac{\tau}{\tau_\infty} \right)^{-3}, \tag{4.3.29}$$

where we suppose that the initial conditions were established at some time $\tau = \tau_\infty$. On the other hand, it is much more useful for comparison with observations (and to be clear about the physical meaning of the solution) to cast this result in terms of cosmic time, for which we need to know the relation between $\tau$ and $t$. This is supplied via the definition of conformal time,

$$\mathrm{d}t = a \, \mathrm{d}\tau \quad \text{so} \quad \mathrm{d}\tau = t^{-2/3} \, \mathrm{d}t. \tag{4.3.30}$$

By direct integration we obtain $\tau \propto t^{1/3}$, so (4.3.26) has solutions in cosmic time which read

$$\delta_c = A(\mathbf{k}) \left( \frac{t}{t_\infty} \right)^{2/3} + B(\mathbf{k}) \left( \frac{t}{t_\infty} \right)^{-1}, \tag{4.3.31}$$

(See Liddle and Lyth (2000, p. 81).) The growing mode is proportional to the scale factor. On superhorizon scales, neglecting pressure in (4.3.22) and picking an appropriate boundary condition gives $\delta_r = (4/3)\delta_c$,[10] whereas on subhorizon scales pressure is dominant, so $\delta_r'' \approx -k^2 \delta_r / 3$ and the radiation perturbation is oscillatory. These oscillations are eventually damped by photon diffusion, a process known as Silk damping.

- Radiation era. Here $\Omega_r \approx 1$ and $\Omega_{\mathrm{CDM}}$ is small. The scale factor goes like $t^{1/2}$, so for the adiabatic mode on superhorizon scales, where we can safely assume that $\delta_r = (4/3)\delta_c$,

$$\delta_c'' + \frac{1}{\tau} \delta_c' - \frac{4}{\tau^2} \delta_c = 0. \tag{4.3.32}$$

As described above, this is an equidimensional equation, so the solutions are of the form $\delta_c \propto \tau^\lambda$, for some $\lambda$ which satisfies the indicial equation $\lambda(\lambda - 1) + \lambda - 4 = 0$. The roots lie at $\lambda = \pm 2$, and to express the result in terms of cosmic time, one only needs the result $a \propto t^{1/2}$, which gives $\tau \propto t^{1/2}$. Therefore the solutions are

$$\delta_c = A(\mathbf{k}) \left( \frac{t}{t_\infty} \right) + B(\mathbf{k}) \left( \frac{t}{t_\infty} \right)^{-1}. \tag{4.3.33}$$

[10] One may add to this any solution of the homogeneous equation $\delta_r'' = 0$, so it clear that this only represents a *particular* solution, the one in which $\delta_r = (2/3)h$ at the initial time, and for which the oscillating mode is absent. It is also the solution for which $h = 2\delta_c$, which was again dependent on the initial conditions. Since there is only *one* initial condition on which this chain of relations depends, it is clear that we have in fact found the adiabatic mode of the perturbation.

On subhorizon scales $\delta_r$ oscillates with average value zero. Therefore, from the trace equation,

$$\delta_c'' + \frac{1}{\tau}\delta_c' = 0. \tag{4.3.34}$$

with solutions

$$\delta_c = A(\mathbf{k}) \ln \tau + B(\mathbf{k}). \tag{4.3.35}$$

The CDM perturbation stagnates and there is no effective growth.

- Vacuum domination. Now $\Omega_\Lambda \approx 1$. The solutions are

$$\delta_c = A(\mathbf{k}) + B(\mathbf{k}) \mathrm{e}^{-2H_0 t} \tag{4.3.36}$$

where $H_0$ is the background Hubble rate. Thus, linear growth ceases.

**4.3.3. The cold dark matter transfer function.** Having studied how the density perturbation $\delta$ grows during each era of interest, we are in a position to project how the primordial fluctuation $\delta(\mathbf{k}, t_\infty)$ evolves from its initial time $t_\infty$ to the present day $t = t_0$. In linear theory there is never any coupling between different $\mathbf{k}$-modes, so each mode evolves independently. Therefore,

$$\delta(\mathbf{k}, t_0) = T(\mathbf{k}) \delta(\mathbf{k}, t_\infty) \tag{4.3.37}$$

where $T(\mathbf{k})$ is called a transfer function, and maps the initial spectrum to what we see at the present epoch. Since it is defined by multiplication in Fourier space, this is evidently a real-space convolution. Recalling (4.3.8), which gives the power spectrum in terms of $\delta$, the present-day power spectrum $\mathcal{P}_0(\mathbf{k})$ is related to the primordial spectrum $\mathcal{P}_\infty(\mathbf{k})$

$$\underbrace{\mathcal{P}_0(\mathbf{k})}_{\text{today}} = T^2(\mathbf{k}) \underbrace{\mathcal{P}_\infty(\mathbf{k})}_{\text{primordial}} \tag{4.3.38}$$

The key changes in the way a perturbation grows depend on when it crosses inside the horizon, from super- to sub-horizon evolution, and when the universe transits from radiation domination in the earliest epochs to matter domination closer to the present day.

(1) Small scales. Small perturbations cross inside the horizon when the universe is still dominated by radiation.

- In the beginning, the fluctuation evolves on superhorizon scales. At some point during this era, it falls within the horizon. It remains subhorizon for the remainder of cosmic history. The growth factor during this era is $(\tau_H/\tau_\infty)^2$, where $\tau_H$ is the conformal time at which horizon crossing occurs.

- The perturbation evolves on subhorizon scales until the onset of matter domination. There is no growth during this era.
- The universe becomes matter dominated. The growth in this phase, until the universe becomes $\Lambda$-dominated, is $(\tau_\Lambda/\tau_{\rm eq})^2$.

Therefore the transfer function is

$$T(\mathbf{k}) = \left(\frac{\tau_H}{\tau_\infty}\right)^2 \left(\frac{\tau_\Lambda}{\tau_{\rm eq}}\right)^2 = \left(\frac{\tau_\Lambda}{\tau_\infty}\right)^2 \left(\frac{\tau_H}{\tau_{\rm eq}}\right)^2 = C\left(\frac{k}{k_{\rm eq}}\right)^{-2}, \tag{4.3.39}$$

where $C$ is a constant, $C = (\tau_\Lambda/\tau_\infty)^2$.

(2) Large scales. Large scale perturbations are still outside the horizon when the universe transits from radiation to matter domination.

- The fluctuation is on superhorizon scales through the matter era. It grows a factor $(\tau_{\rm eq}/\tau_\infty)^2$ during this time.
- The universe becomes matter dominated. Between this event and the perturbation crossing the horizon, it grows by a factor $(\tau_H/\tau_{\rm eq})$.
- The universe is still matter dominated, but the perturbation is subhorizon. It grows until the onset of $\Lambda$-domination, by a factor $(\tau_\Lambda/\tau_H)^2$.

Thus the transfer function is

$$T(\mathbf{k}) = \left(\frac{\tau_{\rm eq}}{\tau_\infty}\right)^2 \left(\frac{\tau_H}{\tau_{\rm eq}}\right)^2 \left(\frac{\tau_\Lambda}{\tau_H}\right)^2 = C. \tag{4.3.40}$$

Therefore, the complete transfer function $T(\mathbf{k})$ is

$$T(\mathbf{k}) = C \times \begin{cases} 1 & k < k_{\rm eq} \\ (k/k_{\rm eq})^{-2} & k > k_{\rm eq} \end{cases} \tag{4.3.41}$$

This simple-minded transfer function is not sophisticated enough to be used in any practical context; instead, one must follow the matter-radiation transition in detail to smooth out the sharp break at $k = k_{\rm eq}$. For a scale invariant power spectrum $\mathcal{P}(\mathbf{k}) = Ak$, the power spectrum we see today is

$$\mathcal{P}_0(\mathbf{k}) = A_0 \times \begin{cases} k/k_{\rm eq} & k < k_{\rm eq} \\ (k/k_{\rm eq})^{-3} & k > k_{\rm eq} \end{cases} \tag{4.3.42}$$

**4.3.4. The variance $\sigma_R$.** Suppose we smooth the density perturbation on a lengthscale $R$ with some window function $W_R(\mathbf{r})$ with characteristic lengthscale $R$. For example, $W_R(\mathbf{r})$ may be a top-hat or Gaussian function. The smoothed density field is

$$\delta_R(\mathbf{r},t) = \int \mathrm{d}^3\mathbf{r}'\, W_R(\mathbf{r}')\delta(\mathbf{r}'-\mathbf{r},t). \tag{4.3.43}$$

Since the mean density perturbation is zero, the variance on a lengthscale $R$ is

$$\sigma_R^2 = \langle \delta_R^2 \rangle. \tag{4.3.44}$$

After rearrangement of the various Fourier transforms appearing here, one can show that

$$\sigma_R^2 = \frac{1}{2\pi^2}\int \frac{\mathrm{d}k}{k}\, k^2 |\widetilde{W_R(k)}|^2 \mathcal{P}(k), \tag{4.3.45}$$

where a tilde denotes the Fourier transform. Setting $W_R$ to include fluctuations in a logarithmic interval around $k$ gives

$$\sigma_R^2 = \frac{k^3}{2\pi^2}\mathcal{P}(k). \tag{4.3.46}$$

Thus, the CDM variance from the projected power spectrum (4.3.42) is today

$$\sigma_R^2 = \langle |\delta_c(\mathbf{k},\tau_0)|^2 \rangle = \frac{AC^2}{2\pi^2} \times \begin{cases} k^4 & k < k_{\mathrm{eq}} \\ k_{\mathrm{eq}}^4 & k > k_{\mathrm{eq}}. \end{cases} \tag{4.3.47}$$

Usually one quotes this quantity evaluated at $8h^{-1}$ Mpc.

## 4.4. The cosmic microwave background

We have previously discussed the evolution of matter and radiation in the universe. The temperature of the universe is determined by the radiation background, from which other species, such as electrons, protons, neutrons and neutrinos decouple as the temperature drops below their mass threshold. Eventually, all species will have decoupled from the radiation which then streams freely through the universe as an essentially non-interacting fluid; this is the cosmic microwave background. The equilibrium processes which were active until decoupling will ensure that this radiation fluid is homogeneous and isotropic, but small perturbations which will inevitably be present on the surface of last scattering will cause temperature fluctuations on the sky. In this section, we describe some causes of these perturbations and calculate the temperature anisotropies to which they give rise.[11]

[11] For a more detailed account, a number of specialist review articles can be consulted, including Durrer (2001b); Hu and Sugiyama (1995).

The simplest case is that of density fluctuations at the surface of last scattering. The temperature variation induced by a density fluctuation must satisfy

$$\delta\rho_r = 4\frac{\pi^2}{15}T^3\delta T \tag{4.4.1}$$

in obedience to the Stefan–Boltzmann law; therefore, $\delta T/T = \delta_r/4$. It is easy to evolve $\delta_r$ forward in time using the results described in the previous section. One finds

$$\frac{\delta T}{T} = \begin{cases} \delta_c(\mathbf{k}, \tau_{\rm dec})/3 & k < k_{\rm dec} \\ -\delta_c(\mathbf{k}, \tau_H)\cos k\tau/\sqrt{3} & k > k_{\rm dec} \end{cases} \tag{4.4.2}$$

where we have used the relationship $\delta_r = (4/3)\delta_c$, and the decoupling wavenumber $k_{\rm dec}$ is defined by $\tau_{\rm dec} = 2\pi/k_{\rm dec}$; the horizon crossing time is $\tau_H = 2\pi/k$. This can be rewritten in terms of the perturbation *today*, as

$$\frac{\delta T}{T} = \frac{1}{3}\delta_c(\mathbf{k}, \tau_0) \times \begin{cases} (\tau_{\rm dec}/\tau_0)^2 & k < k_{\rm dec} \\ (\tau_H/\tau_0)^2\cos k\tau/\sqrt{3} & k > k_{\rm dec} \end{cases} \tag{4.4.3}$$

The present conformal time can be calulated from the relation $a \propto \tau^2$; it is $\tau_0 = 2\mathcal{H}_0^{-1}$, where $\mathcal{H} = a'/a$ is the conformal Hubble rate. The variance in temperature fluctuations is then

$$\left(\frac{\delta T}{T}\right)^2\bigg|_k = \frac{1}{9}\langle|\delta_c(\mathbf{k}, \tau_0)|^2\rangle\pi^4\mathcal{H}^4 \times \begin{cases} k_{\rm dec}^{-4} & k < k_{\rm dec} \\ k^{-4}\cos^2(2\pi k/\sqrt{3}k_{\rm dec}) & k > k_{\rm dec} \end{cases} \tag{4.4.4}$$

One can also consider Doppler effects, which give rise to very simple temperature anisotropies. Velocities in the radiation fluid cause Doppler shifts

$$\frac{\Delta T}{T} = \frac{\Delta\nu}{\nu} = -\Delta v = \hat{\mathbf{n}}\cdot\mathbf{v} \tag{4.4.5}$$

where $\hat{\mathbf{n}}$ is a unit vector in the photon's direction of propagation.

**4.4.1. The Sachs–Wolfe integral.** There will also be fluctuations owing to the gravitational perturbations induced by density perturbations at last scattering, which were first described in detail by Sachs and Wolfe. These dominate the power spectrum at large angles, where processes in the local universe cannot modify primordial anisotropies. There are many essentially equivalent ways to carry out this calculation. One can find at least three different approaches in the standard literature (Liddle and Lyth, 2000; Peacock, 1999), and other clear expositions are available (White and Hu, 1997). We present an

alternative derivation here, based on the synchronous gauge that we have been using until now.[12] To this end, consider photon propagation in a perturbed FRW background, in the synchronous gauge,

$$\mathrm{d}s^2 = a^2(\tau)\left[\mathrm{d}\tau^2 - (\delta_{ij} - h_{ij})\mathrm{d}x^i\,\mathrm{d}x^j\right] = \mathrm{d}t^2 - \mathrm{d}\mathbf{r}^2. \tag{4.4.6}$$

A photon trajectory is described by $\mathrm{d}t = |\mathrm{d}\mathbf{r}| = \mathrm{d}r$. The unperturbed comoving trajectory in the photon propagation direction $\hat{\mathbf{n}}$ is

$$\mathbf{x} = \hat{\mathbf{n}}\tau \tag{4.4.7}$$

which is the zero'th order approximation to the perturbed path; so,

$$\delta x_\alpha = \hat{n}_\alpha \delta\tau. \tag{4.4.8}$$

To find the first order corrections, observe that the proper separation of two comoving observers will be

$$\mathrm{d}r = a\left(1 - \frac{1}{2}h_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\right)\delta\tau, \tag{4.4.9}$$

to first order in $h_{\alpha\beta}$. Therefore, the velocity variation during the time interval $\delta t = \delta r$ is

$$\mathrm{d}v = \frac{\mathrm{d}}{\mathrm{d}t}\delta r \simeq \left[\dot{a}\left(1 - \frac{1}{2}h_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\right) - \frac{1}{2}\dot{h}_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\right]\delta\tau. \tag{4.4.10}$$

One can now eliminate the $\tau$ dependence, by dividing by

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{\mathrm{d}r}{\mathrm{d}\tau} \tag{4.4.11}$$

to get

$$\mathrm{d}v\frac{\mathrm{d}\tau}{\mathrm{d}t} = \left[\dot{a}\left(1 - \frac{1}{2}h_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\right) - \frac{1}{2}\dot{h}_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\right]\mathrm{d}\tau \Big/ a\left(1 - \frac{1}{2}h_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\right). \tag{4.4.12}$$

To first order, this reads

$$\mathrm{d}v = \left(\frac{\dot{a}}{a} - \frac{1}{2}\dot{h}_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\right)\mathrm{d}t. \tag{4.4.13}$$

The $\dot{a}/a$ term is the Hubble flow, which just records the effect of the expansion of the universe. This is a systematic effect, unrelated to the influence of any perturbations in the metric or matter fields. Since systematic effects of this kind are not what we are trying to

[12] The synchronous gauge is not at all the best venue for this calculation.

measure, this term should be subtracted. To relate the result to temperature anisotropies, we integrate from decoupling to the present day,

$$\int_{t_{\rm dec}}^{t_0} \frac{{\rm d}\nu}{\nu} = \ln\frac{\nu_0}{\nu_{\rm dec}} = \frac{1}{2}\int_{t_{\rm dec}}^{t_0} \dot{h}_{\alpha\beta}\hat{n}^\alpha\hat{n}^\beta\,{\rm d}t \tag{4.4.14}$$

The logarithm of the frequency shift can be expanded to first order,

$$\ln\frac{\nu_0}{\nu_{\rm dec}} = \ln\frac{\nu_0}{\nu_0+\Delta\nu} = -\ln\left(1+\frac{\Delta\nu}{\nu_0}\right) \simeq \frac{\Delta\nu}{\nu_0} \simeq \frac{\Delta T}{T}. \tag{4.4.15}$$

This is not usually how the Sachs–Wolfe effect is written (Liddle and Lyth, 2000; Peacock, 1999; White and Hu, 1997), where the effect is related to the Newtonian potential $\phi$. This can be recovered from (4.4.14) by a change of gauge, from the synchronous gauge we have been using so far to the Newtonian gauge, in which the perturbation of the time–time metric component is related to the Newtonian gravitational potential.

To obtain the usual effect (White and Hu, 1997), write the Sachs–Wolfe integral as

$$\frac{\Delta T}{T} = \frac{1}{2}\int {\rm d}\tau \sum_{\mathbf{k}} {\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau}\left[\frac{1}{3}h'\delta_{ij} + \left(\hat{k}_i\hat{k}_j - \frac{1}{3}\delta_{ij}\right)h'_S\right], \tag{4.4.16}$$

where we have decomposed $h_{ij}$ into scalar modes,

$$h_{ij} = \frac{1}{3}h\delta_{ij} + \left(\nabla_i\nabla_j - \frac{1}{3}\delta_{ij}\right)h_S \tag{4.4.17}$$

where $h_S$ does not contribute to the trace. Collecting terms, this is

$$\frac{\Delta T}{T} = \frac{1}{2}\int {\rm d}\tau \sum_{\mathbf{k}} {\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau}\left[\frac{1}{3}(h'-h'_S) + (\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})^2 h'_S\right]. \tag{4.4.18}$$

Evaluating the second term by parts, we obtain

$$\frac{1}{2}\int {\rm d}\tau \sum_{\mathbf{k}} (\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})^2 h'_S {\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau} = -\left[\sum_{\mathbf{k}} \frac{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}}{2k^2}h'_S{\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau}\right]_{\tau_{\rm dec}}^{\tau_0} + \int {\rm d}\tau \sum_{\mathbf{k}} \frac{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}}{2k^2}h''_S{\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}}, \tag{4.4.19}$$

where we have integrated from decoupling to the present day, and integrating by parts a second time yields

$$\frac{1}{2}\int {\rm d}\tau \sum_{\mathbf{k}} (\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})^2 h'_S {\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau} = -\left[\sum_{\mathbf{k}} \frac{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}}{2k^2}h'_S{\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau}\right]_{\tau_{\rm dec}}^{\tau_0} + \left[\sum_{\mathbf{k}} \frac{h''_S}{2k^2}\right]_{\tau_{\rm dec}}^{\tau_0} - \int {\rm d}\tau \sum_{\mathbf{k}} \frac{h'''_S}{2k^2}{\rm e}^{{\rm i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau}. \tag{4.4.20}$$

By massaging this expression a little, we have

$$\frac{\Delta T}{T} = -\left[\sum_{\mathbf{k}} \frac{\mathrm{i}\mathbf{k}\cdot\hat{\mathbf{n}}}{2k^2} h'_S \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau}\right]_{\tau_{\mathrm{dec}}}^{\tau_0} + \left[\sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\hat{\mathbf{n}}} \Big( \underbrace{\frac{h''_S}{2k^2} + \frac{a'}{a}\frac{h'_S}{2k^2}}_{\phi} - \frac{a'}{a}\frac{h'_S}{2k^2} \Big)\right]_{\tau_{\mathrm{dec}}}^{\tau_0} \tag{4.4.21}$$
$$+ \int \mathrm{d}\tau \sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\hat{\mathbf{n}}\tau} \underbrace{\left(\frac{1}{6}(h' - h'_S) - \frac{h'''_S}{2k^2}\right)}_{\phi'+\psi'=2\phi'}.$$

The evaluation of the first bracket at the present day $\tau_0$ contributes a dipole and the evaluation of the second contributes a monopole. Both of these can be ignored. Using the synchronous–Newtonian relations

$$\psi = \frac{1}{6}(h - h_S) + \frac{a'}{a}\frac{h'_S}{2k^2} \tag{4.4.22}$$

$$\phi = -\frac{h''_S}{2k^2} - \frac{a'}{a}\frac{h'_S}{2k^2}, \tag{4.4.23}$$

the total temperature anisotopy, including density perturbations, Doppler effects, and gravitational fluctuations, can be written as

$$\frac{\Delta T}{T} = \frac{1}{4}\delta_r^N + \mathbf{v}^N \cdot \hat{\mathbf{n}} + \frac{1}{3}\phi + 2\int_{\tau_{\mathrm{dec}}}^{\tau_0} \phi' \,\mathrm{d}\tau, \tag{4.4.24}$$

where a superscript $N$ denotes Newtonian gauge quantities. The term $\phi$ denotes the potential perturbation on the surface of last scattering, whereas the integral term is known as the integral Sachs–Wolfe effect.

## 4.5. Inflation

**4.5.1. Introduction.** Although observations indicate that the standard cosmological model represents a good approximation to the true state of the universe, it cannot be a complete theory of cosmology if only because there is no natural explanation in the theory for the degree of large scale flatness and isotropy which are actually seen. Although, as we have described, there is a presumption of homogeneity and isotropy from the outset in the standard model, this applies only to the background. Once perturbations are taken into account, the situation changes dramatically. To see how this works in detail, consider any spatially homogeneous and isotropic FRW model, of curvature $k = \{\pm 1, 0\}$.

There are two separate problems. The first of these is called the flatness problem and relates to the late-time geometry of the universe. Eq. (4.2.8) shows that the evolution of the density parameter $\Omega$ is governed by the rule $\dot{\Omega} = 2Hq(\Omega - 1)$. Assuming $H$ is positive

| | $q > 0$ | $q < 0$ |
|---|---|---|
| $\Omega = 1$ | $\Omega = 1$ | $\Omega = 1$ |
| $\Omega > 1$ | repeller ($\Omega \to \infty$) | attractor ($\Omega \to 1$) |
| $\Omega < 1$ | repeller ($\Omega \to 0$) | attractor ($\Omega \to 1$) |

**Table 1.** Dynamical behaviour of $\Omega$ (see eg., Wainwright and Ellis (1997))

(almost certainly a safe assumption over the lifetime of the universe, although some models contain epochs of recollapse culminating in an infelicitously named Big Crunch which signals the end of the lifetime of the present universe), $\Omega$ behaves as described in Table 1. In particular, the behaviour of $\Omega$ depends crucially on the sign of $q$.[13] For a universe dominated by familiar matter, such as radiation or matter, we have already seen that one obtains scale factors which scale with $t$ like $a \propto t^{2/3}$ or $a \propto t^{1/2}$, or more generally like $t^n$ for $0 < n \leqslant 2/3$. For any such cosmology, $q > 0$.[14] Thus, the separatrix $\Omega = 1$ is an attractor when $q < 0$ but a repeller otherwise. Approximately flat universes with $\Omega \approx 1$ at late times are therefore rather unlikely, provided the universe is filled with normal matter which obeys an equation of state $p = \omega\rho$ with $\omega \geqslant 0$.

Of course, this problem disappears at a stroke if one assumes that $k = 0$, for which $\Omega = 1$ for all time, but the problem here is that such universes are very special, and presumably highly unlikely. Although we have no idea how to specify the initial conditions which determine $k$, it would appear *a priori* to be quite probable that initial conditions leading to $\Omega = 1$ would constitute a set of measure zero among all allowable initial conditions for the universe. Although this statement is not meant to be very precise, since we have no idea what the allowable initial conditions are, in some sense measure zero events are not supposed to occur in physics. Whatever process set the initial conditions *presumably* had

[13] In the early days of cosmology, measuring the sign of the deceleration parameter was one of the goals of observational cosmology; for example, this approach is emphasized in the early monograph by Weinberg (1972). In modern work $q$ appears more seldom, having been notationally superseded by the inflationary parameter $\varepsilon$, to which it is equivalent. The goal is not really to measure $\varepsilon$ (or $q$) for the present universe, but rather in the early universe, where it will tell us something about the particle physics potential presumably supporting inflation. This will be discussed later.

[14] To check that this really is the case, it suffices to recall the redshifting law $\rho \propto a^{-3(1+\omega)}$. The resulting scale factor is proportional to $t^{2/3(1+\omega)}$, so $t^{2/3}$, which occurs where $\omega = 0$, is the strongest scaling with $t$ which can be expected. To have $q < 0$, one must have $\omega < -1/3$. Such matter has negative pressure

only a vanishingly small probability of alighting on $\Omega = 1$ by accident, unless we admit the possibility of some sort of intelligent design[15] in which $\Omega = 1$ was chosen deliberately. Since observations indicate that the universe actually is very close to flat (some 13 Gyr after the big bang), this is a serious problem.

The argument can be made more precise. The relation (4.2.31) shows that, when no species are decoupling, $a \propto T^{-1}$, or simply that $a$ scales as the inverse temperature.[16] Suppose for simplicity that the universe is radiation dominated throughout its history. Then,

$$1 - \Omega = \frac{k}{a^2 H^2} \propto \frac{1}{a^2 \rho} \propto a^2 \propto T^{-2}. \tag{4.5.1}$$

The present microwave background temperature is $T_{\rm CMB} \sim 10^{-13}\,\mathrm{GeV}$,[17] so

$$|\Omega_{\rm P} - 1| \sim 10^{-64} |\Omega_0 - 1| \tag{4.5.2}$$

where $\Omega_{\rm P}$ is the curvature parameter at the Planck epoch. More conservative estimates, taking into account that the universe is not radiation dominated throughout its history, give a slightly weaker figure $|\Omega_{\rm P} - 1| \sim 10^{-60}$ (Peacock, 1999). Therefore the curvature parameter must be very strongly fine-tuned, to the extent of sixty orders of magnitude.

To ameliorate these difficulties, one must find an extension of the standard model with can naturally accommodate the flatness and isotropy of the present universe. One can either seek to append new physics to the present model, or one can seek an entirely new theory (presumably relying on a different formulation of the gravitational and matter dynamics) into which the observational data can be naturally accommodated. For reasons of pragmatism and economy, and also because it is difficult to find viable alternative cosmologies, the focus of research has usually centred around the former option.

Among the various conjectural extensions of the standard model which have been proposed, inflation (Guth, 1981) enjoys a favoured status as the most promising candidate for a solution to the problems described above. Before proceeding, it is necessary to point out that that there is no unique candidate *model* of inflation that arises from particle

[15] In which case, we may as well give up the business of physics and go fishing, because second-guessing whatever hypothetical creator we choose to invoke is going to be a desperate job.

[16] This was derived in (4.2.31) on the basis on entropy arguments, but can be obtained more simply in a relativistically dominated universe just by remembering that radiation scales like $a^{-4}$. Since $\rho_r \propto T^4$, the result follows.

[17] The conversion from Kelvin to eV is, on order of magnitude, $1\,\mathrm{K} = 10^{-4}\,\mathrm{eV}$.

physics, either speculative or well-established – only a scenario. This scenario is remarkably general, since all that is required is that the inflationary condition $q < 0$ is satisfied. Because of this liberality, it is possible to construct inflationary models in field theory and string theory equally, using a selection of field content or geometrical backgrounds (Boyanovsky, Cao, and de Vega, 2002; Garcia-Bellido, Rabadan, and Zamora, 2002; Halyo, 2002b, 2003; Kachru, Kallosh, Linde, Maldacena, McAllister, and Trivedi, 2003a; Kachru, Kallosh, Linde, and Trivedi, 2003b; Lyth and Riotto, 1999). There is a general interpretation of inflation in terms of the renormalization group flow near an ultra-violet fixed point which we will discuss later (Larsen, van der Schaar, and Leigh, 2002) (Section 4.5.3).

Notwithstanding this enormous theoretical flexibility (and despite numerous attempts (Lyth and Riotto, 1999)), it has proved entirely impossible to derive an inflationary epoch from the field content of the Standard Model, and the situation is no better in the minimally supersymmetric Standard Model (usually abbreviated MSSM to minimise clumsy phrasing). Instead, theoretical investigations usually centre on so-called chaotic models of inflation, where typically the field driving inflation is descending from the Planck scale, and only the highest power appearing in its potential is usually relevant.[18] However, no model of this type is really viable or self-consistent, both because inflation must occur at field values above the Planck scale, where field theory is not under good control,[19] and also because in naïve field theory models it is impossible (without unsatisfactory fine tuning) to prevent the inflaton acquiring a renormalized mass of order the largest mass in the theory, in this case presumably the Planck scale. Such a heavy field cannot drive inflation, at least not of the slow-roll variety, because a large mass implies that the curvature of the potential is too great.

For example, consider the Taylor expansion of any general potential $V(\phi)$ around some extremal point $\phi_0$,

$$V(\phi) = V(\phi_0) + \frac{1}{2}V''(\phi_0)(\phi - \phi_0)^2 + \cdots = V(\phi_0) + \frac{b}{2}\frac{V(\phi_0)}{M_{\rm P}^2}(\phi - \phi_0)^2 + \cdots , \qquad (4.5.3)$$

[18]This was not implied in Linde's original formulation of chaotic inflation, where the idea applied only to the initial conditions and no requirement of large field values was made. But in the intervening time, the idea has become synonymous with large-field inflation, and we follow that convention here.

[19]In fact it is easy to see that for a polynomial model where $V \sim \phi^n$, inflation occurs when $\phi > nM_{\rm P}/\sqrt{4\pi}$. In realistic models, inflation should always be occurring when the field has a value less than the Planck scale.

where on dimensional grounds we have written $V'' = bV/M_{\rm P}$, and $b$ is a dimensionless number which would naturally be of order unity. If the gravitational theory is supersymmetric, as would naturally occur in straightforward supergravity or in cosmologies descending from a string theory, then one might hope that the non-renormalization theorems of supersymmetry would protect this potential against *perturbative* renormalization. During an inflationary epoch, $V$ is related to the Hubble parameter, so one has

$$V(\phi) = V(\phi_0) + \frac{1}{2}\left(H\sqrt{\frac{3b}{8\pi}}\right)^2 (\phi - \phi_0)^2 + \cdots \tag{4.5.4}$$

Therefore the field $\phi$ acquires a mass during inflation of order $H$. (In fact, the choice of $M_{\rm P}$ as the dimensionful scale is only really correct at the Planck scale, and for inflation at lower energies one should follow the renormalization group flow for the mass. One should not expect this subtlety to change the situation drastically.) Although this mass is slightly too large to comfortably support inflation, the situation is vastly improved in comparison with the usual case where $\phi$ has a mass of order the Planck scale. One can suppress the mass a little by tuning $b$, but suppression by many orders of magnitude cannot be natural.

**4.5.2. Scalar field cosmology.** The action for a scalar field is

$$S_\phi = -\int \mathrm{d}^4x\, \sqrt{-g}\left(\frac{1}{2}g^{ab}\nabla a\phi \nabla b\phi + V(\phi)\right). \tag{4.5.5}$$

The overall sign is chosen to make the timelike derivative appear with positive coefficient; the Hamiltonian is then positive, up to a possible infinite constant. To find the equation of motion for $\phi$, one uses the fact that the Lagrangian density is equivalent to $\frac{1}{2}\phi\Box\phi - V$, and since $\Box = \nabla^a\nabla_a$ is self-adjoint one has

$$\Box\phi - V' = 0, \quad \text{or, in flat space,} \quad \ddot{\phi} - \triangle\phi = -V' \tag{4.5.6}$$

where a prime $'$ denotes a derivative with respect to $\phi$. The sign of the potential $V$ in the action is fixed by observing that if $V$ is a mass term, $V = \frac{1}{2}m^2\phi^2$ then this reduces to

$$\ddot{\phi} - \triangle\phi = -m^2\phi. \tag{4.5.7}$$

The Killing vectors of flat space are

$$E \sim -\mathrm{i}\frac{\partial}{\partial t} \quad \text{and} \quad p_\alpha \sim \mathrm{i}\frac{\partial}{\partial x^\alpha}, \tag{4.5.8}$$

so (4.5.7) merely enforces the relativistic energy–momentum relation $E^2 = \mathbf{p}^2 + m^2$ as an operator equation, which is exactly what we want.

The energy–momentum tensor can be recovered from Noether's theorem, but for coupling to gravity it is much more useful to adopt the variational definition

$$T_{ab} = -\frac{2}{\sqrt{-g}}\frac{\delta I_\phi}{\delta g^{ab}}, \tag{4.5.9}$$

and after a straightforward calculation one finds

$$T^a{}_b = \nabla^a\phi\nabla_a\phi - \frac{1}{2}\delta^a_b g^{cd}\nabla_c\phi\nabla_d\phi - V\delta^a_b \tag{4.5.10}$$

To extract the density and pressure of the $\phi$-fluid, one can compare with the energy–momentum tensor of a perfect fluid, $T^a{}_b = \mathrm{diag}(-\rho, p, p, p)$, so in flat space the density and isotropic average pressure satisfy

$$\rho_\phi = \frac{1}{2}\dot\phi^2 + \frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + V \quad \text{and} \quad p_\phi = \frac{1}{2}\dot\phi^2 - \frac{1}{6}(\boldsymbol{\nabla}\phi)^2 - V. \tag{4.5.11}$$

Nevertheless, the scalar field has in general no equation of state. On the other hand, if $\boldsymbol{\nabla}\phi = 0$, so that spatial gradients vanish and the kinetic term $\dot\phi/2$ can be neglected in comparison with the potential $V$, then $p_\phi \approx -\rho_\phi$, approximating matter with equation of state parametrized by $\omega = -1$. As discussed above, $\omega < -1/3$ is a requirement for inflation. Neglecting spatial gradients is equivalent to insisting that the field configuration $\phi$ be homogeneous.

The same expressions hold with a Robertson–Walker line element provided that the spatial gradients are suppressed by $1/a^2$, so since the field must be homogeneous, the Friedmann equation for $a$ in a universe consisting solely of a scalar field becomes

$$H^2 = \frac{8\pi}{3M_{\mathrm{P}}^2}\left(\frac{1}{2}\dot\phi^2 + V\right). \tag{4.5.12}$$

Evidently, if $\phi$ does not move much in a Hubble time then $V$ will support a quasi-de Sitter epoch. This Friedmann equation can be recast as a form of the $\phi$ equation of motion,

$$H'(\phi) = -\frac{4\pi}{M_{\mathrm{P}}^2}\dot\phi(t) \tag{4.5.13}$$

where a prime $'$ denotes a derivative with respect to $\phi$ and an overdot denotes a derivative with respect to the cosmic time $t$. In practice, this equation allows one to trade $t$ derivatives for $\phi$ derivatives, and *vice-versa*.

One defines an inflationary epoch by the condition $\ddot a > 0$, or $q < 0$. Reference to Table 1 shows that $\Omega$ is an attractor during inflation, so the universe is dynamically driven to $\Omega = 1$ and spatial flatness. An epoch of inflation, provided it is sufficiently prolonged,

reverses the situation: instead of being catastrophically unlikely, as indicated by (4.5.2), late-time flat universes become rather natural. Since $a$ is positive,

$$\text{inflation implies} \quad 0 < \frac{\ddot{a}}{a} = \dot{H} + H^2 \quad \text{or,} \quad \varepsilon \doteq -\frac{\dot{H}}{H^2} < 1, \tag{4.5.14}$$

so $\varepsilon < 1$ is a necessary and sufficient condition for inflation. As we noticed before, $\varepsilon$ is related to the deceleration parameter $q$ by $q = \varepsilon - 1$. The parameter $\varepsilon$ turns out to be exceedingly convenient and constitutes one of a series of parameters, known as slow-roll parameters, which characterize the inflationary epoch. Application of (4.5.13) allows this to be recast without reference to $t$,

$$\varepsilon = \frac{M_{\mathrm{P}}^2}{4\pi}\left(\frac{H'}{H}\right)^2. \tag{4.5.15}$$

Define a physical length scale $\ell_H^{\mathrm{phys}}$, called the Hubble length,[20] by setting $\ell_H^{\mathrm{phys}} = H^{-1}$. The comoving length scale corresponding to the Hubble length is $\ell_H^{\mathrm{com}} = 1/aH$, which changes with time according to

$$\dot{\ell}_H^{\mathrm{com}} = -\frac{1}{a^2 H}\dot{a} - \frac{\dot{H}}{aH^2} = \frac{\varepsilon - 1}{a}. \tag{4.5.16}$$

Therefore the comoving Hubble length shrinks during inflation, whereas a quantity of fixed comoving scale grows relative to $\ell_H^{\mathrm{com}}$. Such scales are rapidly inflated outside the causal horizon. A common parametrization of the duration of any inflationary epoch is to measure the change in the comoving Hubble length,

$$\text{duration} = \frac{\ell_H^{\mathrm{com}}|_{\mathrm{start}}}{\ell_H^{\mathrm{com}}|_{\mathrm{end}}} = \frac{a_{\mathrm{end}} H_{\mathrm{end}}}{a_{\mathrm{start}} H_{\mathrm{start}}}. \tag{4.5.17}$$

[20]The terms *Hubble length* and *horizon* are typically used interchangeably is cosmology, although in principle one should take care to preserve the distinction. The local Hubble law relating the recession velocity $v$ of some distant object to its distance $d$ is $v = Hd$, so the Hubble length is the distance at which the naïve Hubble law predicts a recession velocity equal to that of light. (At this distance, one needs to take higher order corrections to $v = Hd$ into account.) In natural units this is numerically equal to the Hubble time, which would be the age of the universe had it always expanded at a rate $H$. The particle horizon is the distance beyond which information about events has not causally had time to reach us, whereas an event horizon (such as the de Sitter horizon) is the distance beyond which information about events will never have time to causally reach us. Usually when cosmologists discuss horizons, they mean particle horizons; event horizons are a global concept and one needs details about the entire future evolution of spacetime to predict where event horizons will fall. Typically the particle horizon and the Hubble length differ by factors of order unity. In the Einstein–de Sitter model, the particle horizon is twice the Hubble length. In de Sitter space, all three horizons coincide.

Typically this is a very large quantity, so its logarithm is used instead to define the number $N$ of e-foldings,

$$N = \ln \text{duration} \quad \text{so,} \quad \mathrm{d}N = \frac{\mathrm{d}a}{a} + \frac{\mathrm{d}H}{H}. \tag{4.5.18}$$

$H$ does not change much during an inflationary epoch, so it is a fair approximation to suppose that the change in $N$ is dominated by $\mathrm{d}a$, since $a$ is – usually – changing exponentially fast with respect to $H$. Thus,

$$\mathrm{d}N \simeq H\,\mathrm{d}t = -\frac{2M_{\mathrm{P}}^2}{8\pi}\frac{H}{H'}\,\mathrm{d}\phi. \tag{4.5.19}$$

One then finds $N$ explicitly by quadrature. As we discussed above, it is necessary to have $|\Omega_{\mathrm{P}} - 1| \lesssim 10^{-60}$ at the Planck epoch in order to accommodate the observed degree of spatial flatness. Since during inflation $H$ moves only a little in comparison with $a$, one has (between times $t_0$ and $t_1$),

$$\mathrm{e}^{2N(t_0 \to t_1)} = \left(\frac{a_1}{a_0}\right)^2 = \frac{1-\Omega_0}{1-\Omega_1}. \tag{4.5.20}$$

If $\Omega$ begins with $\Omega = O(1)$ at time $t_0$, then at time $t_1$ the necessary fine-tuning will have been achieved for $N \approx 70$ e-folds. At $t_0$, the Hubble volume is supposed by some means to have become causally homogenized, so our present Hubble volume will still be homogeneous provided super-Hubble scales at $t = t_0$ are still super-Hubble today. In the radiation domination approximation, there have been $N \approx \ln(T_1/T_{\text{today}})$ e-folds between the end of inflation at time $t = t_1$ and today, which gives $N \approx 70$ for $t_1$ at the Planck epoch. This equality between the number of e-folds needed to solve the flatness and horizon problems is generic, irrespective of the energy scale of inflation (Peacock, 1999).

**4.5.3. Inflation and transplanckian physics.** Despite its many attractive features, it necessary to guard against any inclination to suppose that inflation solves or removes any of the problems of the Standard Model. It is at best a mitigating influence, which comes complete with problems of its own. Apart from the overall difficulty of constructing inflation within the framework of physics as it exists today (that is, a gauge field theory with $SU(3)\times SU(2)\times U(1)$ gauge group and fermionic matter, together with a minimally coupled gravitational sector), there are also difficulties with the scenario as a whole (Hollands and Wald, 2002). Among the most serious of these is the so-called transplanckian problem (Brandenberger, 2002; Martin and Brandenberger, 2002), which consists in the observation that the enormous stretching of physical lengthscales that solves the horizon and flatness

problems also pushes strongly ultra-violet oscillations at or above the Planck scale onto astronomical lengthscales. Since Planck-scale physics is not under good control, there is no reason to believe that we understand the quantum vacuum there. Consequently it is difficult to unambiguously predict exactly what characteristics we should see in the oscillation when it is expanded to an astronomical lengthscale.[21] One can attempt to phenomenologically describe the effect of transplanckian physics by modifying or replacing the traditional dispersion relation $E^2 = \mathbf{p}^2 + m^2$ and calculating the effect on CMB spectra (Martin and Brandenberger, 2002).

If such effects could be measured in the power spectrum of the cosmic microwave background, or other cosmological fossils from an epoch of the universe when transplanckian scales could be probed, then observations could tell us a great deal about the phenomenological quantum vacuum above the Planck scale. Although there is still no microphysical model for this vacuum, there is no reason why this should hinder a phenomenological parametrization, and indeed such a description might offer valuable insight into the degrees of freedom supplied by a full quantum gravity. Working in reverse, one can use speculative ideas about the nature of quantum gravity to probe for potentially observable effects in the cosmic microwave background. For example, if the entropy in any spacetime volume is bounded by the entropy of the boundary, as is expected in any theory of gravity obeying the holographic principle (Susskind, 1995), then the spectrum of perturbations may be discrete (Hogan, 2002).

## 4.6. Perturbations and the origin of structure

Inflation was not originally cast in these terms, since it was designed to dilute a population of monopoles that might have been been produced when breaking a GUT gauge group to the standard model $SU(3) \times SU(2) \times U(1)$. In the process inflation can make natural the astoundingly large entropy of the universe (Blau and Guth, 1987). However, it was quickly noticed that inflation had a beneficial unintended side-effect, in the sense that it could provide a mechanism to lay down perturbations in the early universe which were of the form (4.3.9) with $n = 1$. Such a 'scale invariant' spectrum had already been proposed by Harrison & Zel'dovich and by Peebles on naturalness grounds.[22]

[21] This ambiguity afflicts Hawking radiation equally (Helfer, 2003).

[22] Essentially the argument is that there are sensible bounds on $n$ that prevent extreme choices (Peacock, 1999). A high value puts too much power on large scales and destroys the FRW background, whereas a low

So far, the treatment of inflation has been entirely classical. We should properly be considering the inflaton as a quantum field, which will be subject to a zero-point fluctuation. As the universe expands, the quantum fluctuation is subject to the same stretching as any other perturbation mode of the universe, meaning (roughly speaking) that oscillations far in the ultra-violet are stretched onto large scales where they freeze in, leading to a characteristic spectrum of curvature fluctuations produced during inflation. In this section, we describe the detailed microphysics leading the this prediction.

**4.6.1. The gauge-invariant description of perturbations.** The line element of an FRW space which has suffered an arbitrary scalar deformation has the general form

$$\mathrm{d}s^2 = -(1+2A)\,\mathrm{d}t^2 + 2a^2\hat{\nabla}_i B\,\mathrm{d}x^i\mathrm{d}t + a^2\left[(1-2\psi)\gamma_{ij} + 2\hat{\nabla}_i\hat{\nabla}_j E\right]\mathrm{d}x^i\mathrm{d}x^j, \qquad (4.6.1)$$

where $\hat{\nabla}$ is the connexion compatible with the three-dimensional, maximally symmetric Euclidean metric $\gamma_{ij}$. Unlike (4.3.10)–(4.3.11), this is in an arbitrary gauge, but we are only including perturbation modes which can be reduced to a description by spin zero fields (Weinberg, 1994) with respect to rotations of the purely spatial variables. Consider a spatial hypersurface described by $\mathrm{d}t = 0$. The induced curvature on this hypersurface satisfies (Wands, Malik, Lyth, and Liddle, 2000)

$$\overset{3}{R} = \frac{6k}{a^2} + \frac{12k}{a^2}\psi + \frac{4}{a^2}\hat{\triangle}\psi, \qquad (4.6.2)$$

where $\hat{\triangle}$ is the $\gamma_{ij}$ Laplacian, and the quantity $\psi$ is defined as the intrinsic curvature perturbation of these hypersurfaces.

The coordinate $t$ usually describes a collection of coincident foliations of spacetime which are described by physically meaningful 'clocks', such as the density or pressure in the universe. The attraction of working in terms of hypersurfaces defined by physical clocks rather than the arbitrary coordinate clock $t$ is that any quantities one recovers from the formalism must be gauge invariant, since the description is framed throughout in terms of observable quantities. The attraction is much the same as the advantage of the Feynman rules in maintaining manifest Lorentz invariance in perturbative calculations in QFT. For example, in the unperturbed case, the hypersurfaces $\mathrm{d}t$ can be described as

---

value produces too much small scale power, overproducing small black holes in the early universe whose consequences should still be visible today.

uniform curvature hypersurfaces, uniform density hypersurfaces, or comoving hypersurfaces. Once one has introduced perturbations, these physically meaningful hypersurfaces do not necessarily coincide. Therefore one must choose a *particular* physical clock with which to measure perturbations. This approach should be contrasted with the gauge-*fixed* coordinate calculations carried out in Section 4.3.2.

Under a gauge transformation $t \mapsto t' = t + \delta t$, it is a simple calculation to show that the intrinsic curvature perturbation transforms according to (Langlois, 2004; Riotto, 2002; Wands et al., 2000)

$$\psi \mapsto \psi' = \psi + H\delta t, \tag{4.6.3}$$

whereas any scalar field such as the density $\rho$, the pressure $P$, or the inflaton field $\phi$ transforms according to the law

$$\rho \mapsto \rho' = \rho - \dot{\rho}\delta t. \tag{4.6.4}$$

Consider the uniform density foliation, where $\rho = 0$ on each spatial slice $\Sigma_\rho$. The intrinsic curvature perturbation on uniform density hypersurfaces must satisfy

$$\psi_\rho = \psi + H\frac{\delta\rho}{\dot{\rho}} \hat{=} -\zeta \tag{4.6.5}$$

where we denote uniform density quantities with a subscript $\rho$. In modern work, it is conventional to denote the intrinsic curvature perturbation on uniform density slices by $\zeta$. As was explained above, $\zeta$ must be a gauge-invariant measure of perturbations, since it is defined in terms of observational quantities, and the same is true for $\psi$ measured on any physically meaningful hypersurface. Another common choice frequently employed in the literature is the curvature perturbation on comoving hypersurfaces,[23] defined as

$$\mathcal{R} = \psi_{\delta\phi=0} = \psi + H\frac{\delta\phi}{\dot{\phi}}. \tag{4.6.6}$$

We will use $\zeta$ exclusively for the description of perturbations during inflation. Occasionally this measure of the curvature perturbation can become ill-defined if the density is not

[23]These are, by definition, hypersurfaces orthogonal to the worldlines of freely falling (or comoving) observers. Since this definition is a little unwieldy, it is useful to know that for practical purposes, comoving observers are those who observe the expansion of the universe to be isotropic, so they see no net momentum flux in their rest frame. During inflation this means that comoving observers are those who see no perturbation in the inflaton field, because the energy–momentum tensor satisfies $T_{0i} \propto \dot{\phi}\partial_i\phi$, so demanding that $T_{0i} = 0$ is equivalent to $\delta\phi = 0$. Therefore comoving hypersurfaces are really the same as uniform-$\phi$ hypersurfaces.

varying monotonically. In this case one can make a different choice of spatial hypersurface, for example by measuring the perturbations via $\mathcal{R}$. We shall not have need of such refinements when calculating the spectrum of perturbations, although the quantity $\mathcal{R}$ itself is frequently useful.

If the hypersurfaces of constant pressure and constant density coincide then the perturbation is said to be adiabatic, and the pressure perturbation is related to the density perturbation by the rule $\delta p = c_s^2 \delta\rho$, where $c_s^2 = \mathrm{d}p/\mathrm{d}\rho = \dot{p}/\dot{\rho}$ is the sound speed of the perturbation. More generally, there will be independent contributions to $\delta p$, in which case one has (Gordon, Wands, Bassett, and Maartens, 2001; Wands et al., 2000)

$$\delta p = c_s^2 \delta\rho + \dot{p}\left(\frac{\delta p}{\dot{p}} - \frac{\delta\rho}{\dot{\rho}}\right) = c_s^2 \delta\rho + \dot{p}\Gamma = c_s^2 \delta\rho + \delta p_{\text{non-adiabatic}}. \tag{4.6.7}$$

This equation contains no physics; it is merely an identity which defines the quantity $\Gamma$. One usually refers to $\Gamma$ as the entropy perturbation or isocurvature perturbation.[24] On large scales where spatial gradients may be neglected, $\zeta$ satisfies the evolution equation (Wands et al., 2000)

$$\dot{\zeta} = -\frac{H}{\rho + p}\delta p_{\text{non-adiabatic}}, \tag{4.6.8}$$

and is therefore constant on large scales in the absence of an isocurvature perturbation. (However, see Gotz (1998) for a comparatively recent discussion of the subtlety of this issue. The constancy of $\zeta$ and other gauge-invariant variables on superhorizon scales can be extended to second order (Vernizzi, 2004), and even a non-perturbative definition can be given (Lyth, Malik, and Sasaki, 2004).)

[24] There are many other ways to describe the decomposition into adiabatic and non-adiabatic terms, but this method seems most convenient. For example, if there is only one matter component, then the perturbation is necessarily adiabatic *provided* the matter has a well-defined equation of state. (If not, for example, a scalar field with spatial gradients, then the perturbation may well be non-adiabatic.) Where two or more fluids are present one can define an isocurvature perturbation as a set of perturbations which sum to zero (Liddle and Lyth, 2000).

The *entropy* terminology is now deprecated, and the more modern term *isocurvature* is preferred in modern work. Despite the terminology, there *is* a perturbation to the space-time metric for isocurvature modes – it enters through the pressure term in the Einstein equations. The curvature referred to is the curvature of comoving hypersurfaces: under an isocurvature perturbation, these hypersurfaces receive no curvature perturbation (Liddle and Lyth, 2000).

**4.6.2. Perturbations from scalar field inflation.** Now suppose the universe contains an evolving scalar field $\phi$, divided into a homogeneous, monotonically rolling classical component $\phi_0$ and a quantum perturbation $\delta\phi$ defined on slices of constant $t$. The perturbation in the scalar field on uniform curvature hypersurfaces is written $Q$ and related to $\delta\phi$ via (4.6.4),

$$Q = \delta\phi + \psi\frac{\dot{\phi}}{H}. \tag{4.6.9}$$

In terms of $Q$ the curvature perturbation on a uniform $\phi$-slicing satisfies

$$\mathcal{R} = \psi + H\frac{\delta\phi}{\dot{\phi}} = \frac{H}{\dot{\phi}}Q, \tag{4.6.10}$$

where we have recalled that the curvature perturbation on comoving hypersurfaces is $\mathcal{R}$. The scalar field perturbation $Q$ defined on flat hypersurfaces is a gauge invariant quantity by definition, and is often called the Mukhanov–Sasaki variable (Mukhanov, 1985; Sasaki, 1986). The density and pressure of the scalar field are determined by (4.5.11), so the perturbations in a quite arbitrary gauge are

$$\begin{aligned}\delta p &= \frac{1}{2}\dot{\phi}\delta(\dot{\phi}) + V'\delta\phi = \frac{1}{2}\dot{\phi}\delta\dot{\phi} - A\dot{\phi}^2 + V'\delta\phi \\ \delta\rho &= \frac{1}{2}\dot{\phi}\delta\dot{\phi} - A\dot{\phi}^2 - V'\delta\phi,\end{aligned} \tag{4.6.11}$$

which is true since one must take account of gauge transformations in the time derivative,

$$\delta(\dot{\phi}) = \delta\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\mathrm{d}(\delta\phi)}{\mathrm{d}t} - \frac{\mathrm{d}\phi}{\mathrm{d}t}\frac{\mathrm{d}\delta t}{\mathrm{d}t} = \delta\dot{\phi} - 2A\dot{\phi}. \tag{4.6.12}$$

Therefore $\delta\rho - \delta p = 2V'\delta\phi$, and since for adiabatic perturbations in the uniform density foliation $\delta\rho$ and $\delta p$ are related via $\delta\rho - \delta p = 0$, the scalar field perturbation $\delta\phi$ also vanishes on uniform density hypersurfaces unless $V' = 0$, which is not generically the case. Thus the comoving and the uniform-density foliation coincide, provided the initial conditions are chosen so that isocurvature modes are absent, which is certainly the case in the minimal model. Therefore, since $-\zeta = \mathcal{R}$, we have the simple relation

$$\zeta = -\frac{H}{\dot{\phi}}Q = -\psi - \frac{H}{\dot{\phi}}\delta\phi. \tag{4.6.13}$$

This argument was first spelt out in detail by Wands et al. (2000).

In a semiclassical approximation, the action consists of Einstein–Hilbert and $\phi$ terms which are treated semiclassically, and expansions of both these terms around their saddle points which are to be treated quantum mechanically. In the early days, estimates of the quantum fluctuations produced by an epoch of inflation typically included only fluctuations

from the inflaton field $\phi$ itself, and discarded metric fluctuations. However, since (as we have seen) the gauge invariant description – in terms of, for example, the invariant-by-construction variables $\zeta$ or $\mathcal{R}$ – *necessarily* couples the perturbations $\delta\phi$ and $\psi$, it is only possible to give a consistent discussion by including both effects. This consistent treatment was begun in the mid-1980s (Mukhanov, 1985, 1998; Sasaki, 1986), working first in terms of the Einstein equations and later in terms of the action itself. The results were finally systematized and assembled in a large-scale review article (Mukhanov et al., 1992) which appeared in the early 1990s, more than ten years after inflation was first proposed.

The theory of coupled metric/inflaton fluctuations is the cornerstone of the theoretical effort to confront early universe theory with observation, and as such its importance can hardly be over-exaggerated. Nor is this all: the calculation of the spectrum of density perturbations produced during inflation is probably the only example of quantum field theory in the presence of external fields which is accessible to observation in the near future. As a result, the study of relic fluctuations potentially has much to teach us about quantum field theory in curved spacetime generally, but unfortunately, most elementary presentations of inflation omit the details of this lengthy calculation. For this reason we pause briefly to outline the most significant steps in the derivation.

Some aspects of Mukhanov's calculation can be gleaned from review articles (Brandenberger, 2002; Riotto, 2002), or the original literature (Mukhanov, 1985, 1998; Sasaki, 1986). However, Mukhanov's original calculation of the fully gauge-invariant result is very complicated, and relied on an expansion of the Einstein–Hilbert action into the degrees of freedom $A$, $B$, $\psi$ and $E$ described by (4.6.1). This expanded action can be simplified via constraint equations and the background equations of motion, before finally being rewritten in terms of gauge-invariant combinations, but the calculation is very long. In this section, we describe a modern simplification due to Maldacena (Maldacena, 2003a).

Before calculating the relevant action, we can begin at the level of the field equations. This provides an outline understanding of what the final result will look like. Following Mukhanov (1985), we work in the conformal Newtonian gauge, where the metric takes the form

$$\mathrm{d}s^2 = a(\tau)^2 \left[-(1+2\psi)\,\mathrm{d}\tau^2 + (1-2\psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right], \tag{4.6.14}$$

and the Einstein equations demand that (being, respectively, the 00, $0i$, and $ij$ equations)[25]

$$2\triangle\psi - 6\mathcal{H}\psi = \kappa^2(2a^2\psi V + a^2\pi V' + \phi'\pi') \tag{4.6.15a}$$

$$2\mathcal{H}\psi + 2\psi' = \kappa^2\phi'\pi' \tag{4.6.15b}$$

$$-4\mathcal{H}^2\psi + 8\frac{a''}{a}\psi + 6\mathcal{H}\psi' + 2\psi'' = \kappa^2(2a^2\psi V - 2\psi\phi'^2 - a^2\pi V' + \phi'\pi'). \tag{4.6.15c}$$

Here we are writing $\pi = \delta\phi$, a prime $'$ denotes a derivative with respect to conformal time $\tau$, and $\mathcal{H} = a'/a$. The second equation is a constraint that relates $\pi$ and $\psi$. Fortunately, it is possible to deal only with the three Einstein equations given above: it is unnecessary to include the scalar field equation, because this arises from energy–momentum conservation $\nabla^a T_{ab} = 0$, and therefore is implicit in the Einstein equations. This is a considerable simplification, because it means that terms involving $\pi''$ do not appear, but only $\pi'$. Subtracting the first equation from the third, using the constraint equation to remove $\pi$, using the background equation

$$\mathcal{H}' - \mathcal{H}^2 = -\kappa^2\phi'^2 \tag{4.6.16}$$

(which can be found by adding the Friedmann equation and the Raychaudhuri equation) to replace $\phi'^2$, and the background scalar field equation $\phi'' + 2\mathcal{H}\phi' + a^2V' = 0$ to eliminate $V'$ gives

$$\psi'' + 2\psi'\left(\mathcal{H} - \frac{\phi''}{\phi'}\right) + 2\psi\left(\mathcal{H}' - \mathcal{H}\frac{\phi''}{\phi'}\right) - \triangle\psi = 0. \tag{4.6.17}$$

One makes the change of variable $\psi = \phi' v/a$ to eliminate the $\psi'$ term. The result is

$$v'' - \triangle v + \left(\mathcal{H}' - \mathcal{H}^2 + \frac{\phi'''}{\phi'} - 2\frac{\phi''}{\phi'}\frac{\phi''}{\phi'}\right)v = 0. \tag{4.6.18}$$

This looks like the theory of a scalar field $v$ in flat space, with a mass term involving a complicated combination of $\mathcal{H}$ and $\phi$. The aim is now to re-write this mass term as something a little simpler. By differentiating (4.6.16), one obtains a simple expression for $\mathcal{H}''$,

$$\mathcal{H}'' = 2\mathcal{H}\mathcal{H}' + 2(\mathcal{H}' - \mathcal{H}^2)\frac{\phi''}{\phi'}. \tag{4.6.19}$$

This lets us rewrite the $v$-equation as the Mukhanov equation (Hwang and Noh, 2002; Mukhanov, 1985),

$$v'' - \left(\triangle + \frac{w''}{w}\right)v = 0, \tag{4.6.20}$$

[25]These equations appeared in a slightly different form in Mukhanov (1985).

where $w$ is defined by

$$w = \frac{\mathcal{H}}{a\phi'} = \frac{H}{a\dot{\phi}}. \tag{4.6.21}$$

The present equation is entirely adequate if all that is required is a solution for $\psi$. On the other hand, if one is aiming for one of the gauge-invariant measures such as $\zeta$ or $\mathcal{R}$, then in the present case the constraint equation shows that

$$-\zeta = \mathcal{R} = \psi + \frac{\mathcal{H}}{\phi'}\frac{2\mathcal{H}\psi + 2\psi'}{\kappa^2\phi'} = \psi + \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'}\frac{1}{a}(a\psi)', \tag{4.6.22}$$

and therefore using (4.6.20) to obtain an expression for $\zeta$ will result in a *third*-order equation in $\psi$. This is inconvenient.[26] In fact, making the transition (Gotz, 1998; Hwang and Noh, 2002), one obtains (in Fourier space)

$$\left(u'' + \mathbf{k}^2 u - \frac{z''}{z}u\right)' - \frac{z'}{z}\left(u'' + \mathbf{k}^2 u - \frac{z''}{z}u\right) = 0 \tag{4.6.23}$$

where $u = -z\zeta$ and $z = w^{-1}$.

To do better, we return to the calculation in terms of the action itself, which in any case is necessary in order to correctly normalize any quantum treatment of the field $\mathcal{R} = -\zeta$. Consider the perturbed theory on flat spatial slices. If we take the description of the relevant degrees of freedom to be a free, massless scalar field, then the action is approximately

$$I^{(0)} = \int \mathrm{d}\tau\, \mathrm{d}^3x\, \frac{a^2}{2}\left(Q'^2 - Q_{,i}Q_{,i}\right), \tag{4.6.24}$$

where $Q$ is the Mukhanov–Sasaki variable and the gauge-invariant curvature perturbation is $\zeta = Q\mathcal{H}/\phi'$. In slow-roll, the background quantities are only slowly varying as a function of time, so that $\mathcal{H}$ and $\phi'$ are not changing rapidly, so the time derivatives of $\phi'/\mathcal{H}$ will be negligible in comparison with derivatives of the field $Q$, which (since it is a quantum fluctuation) is supposed to be undergoing rapid oscillation, so the action becomes

$$I^{(0)} = \int \mathrm{d}\tau\, \mathrm{d}^3x\, \frac{z^2}{2}\left(\zeta'^2 - \zeta_{,i}\zeta_{,i}\right), \quad \text{where} \quad z = \frac{a\phi'}{\mathcal{H}}. \tag{4.6.25}$$

This argument gives the correct slow-roll form, but in fact (4.6.25) is an exact expression for the action. To see this, consider that under a gauge transformation $x^a \mapsto \tilde{x}^a = x^a + \delta x^a$,

[26] However there is no reason why one *cannot* solve for $\psi$ and then use the present equation to reconstruct $\zeta$. In effect, this is what was done in Sasaki (1986).

the perturbations in the metric (4.6.1) transform according to the laws

$$\tilde{A} = A - \xi^{0\prime} - \frac{a'}{a}\xi^0 \tag{4.6.26a}$$

$$\tilde{B} = B + \xi^0 + \beta' \tag{4.6.26b}$$

$$\tilde{\psi} = \psi - \frac{1}{3}\triangle\beta + \frac{a'}{a}\xi^0 \tag{4.6.26c}$$

$$\tilde{E} = E + 2\beta \tag{4.6.26d}$$

where we have parametrized $\delta x^a = (\xi^0, \nabla_i\beta)$. When transforming to flat spatial slices, we inevitably pick up a deformation in $A$ and $B$, so it is not correct to approximate the action by the flat space form (4.6.24); instead, there will be non-trivial contributions from the $g_{00}$ and $g_{0i}$ fields in the metric. Mukhanov's original calculation of the exact form of (4.6.25) is so complicated because these extra metric fields are handled by brute force. Instead, it is more convenient to use the ADM description of the metric (Maldacena, 2003a), in which

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + h_{ij}(\mathrm{d}x^i + N^i\mathrm{d}t)(\mathrm{d}x^j + N^j\mathrm{d}t). \tag{4.6.27}$$

The gravitational action has a well-known form in terms of these variables (D'Eath, 1996),

$$R = \overset{3}{R} - (\mathrm{Tr}\,K)^2 + \mathrm{Tr}\,K^2 + \text{total derivative} \tag{4.6.28}$$

where the total derivative can be ignored when calculating the action, and the extrinsic curvature $K_{ij}$ satisfies

$$K_{ij} = \frac{1}{N}\left(\frac{1}{2}\dot{h}_{ij} - N_{(i|j)}\right) = \frac{1}{N}E_{ij}. \tag{4.6.29}$$

The derivative $|$ is the covariant derivative compatible with $h_{ij}$. The the coupled Einstein/scalar field system has action

$$I^{(0)} = \int \mathrm{d}t\,\mathrm{d}^3x\,\frac{1}{2}\sqrt{h}\left[\frac{N\overset{3}{R}}{\kappa^2} + \frac{1}{N\kappa^2}(E_{ij}E^{ij} - E^2) + N(\dot{\phi} - N^i\phi_{,i})^2 - Nh^{ij}\phi_{,i}\phi_{,j} - 2NV\right], \tag{4.6.30}$$

where indices $i$, $j$, ... are raised and lowered with $h_{ij}$. The advantage of this representation is that the fields $N$ and $N^i$, which encode the $g_{00}$ and $g_{0i}$ fields of the metric and therefore contain the gauge-corrections after transforming to flat spatial slices, are known to be Lagrange multipliers (D'Eath, 1996). For this reason, their equations of motion are entirely algebraic, and once solved can be back-substituted into the action to obtain a correct description of the reduced degrees of freedom; this process was described in Chapter 2. Moreover, it is only necessary to solve $N$ and $N^i$ to first order in the perturbations, since

any second order terms in these fields must multiply the $\partial N$ derivative of the background action, evaluated to zero'th order. This vanishes, since the $N$ field equation is

$$\frac{\partial L}{\partial N} = 0, \tag{4.6.31}$$

where $L$ is the integrand in (4.6.30). A similar argument can be made for $N^i$. The $N$ equation of motion, or constraint, is

$$\frac{\overset{3}{R}}{\kappa^2} - \frac{1}{N^2\kappa^2}(E^{ij}E_{ij} - E^2) - \frac{1}{N^2}(\dot{\phi} - N^i\phi_{,i})^2 - h^{ij}\phi_{,i}\phi_{,j} - 2V = 0. \tag{4.6.32}$$

It turns out that the $N^i$ equation is not needed. We take

$$N = 1 + \alpha \quad \text{and} \quad N_i = \partial_i\psi, \tag{4.6.33}$$

and pick the gauge

$$\delta\phi = 0 \quad \text{and} \quad h_{ij} = a^2(1 + 2\mathcal{R})\delta_{ij} \tag{4.6.34}$$

where $t$ parametrizes comoving hypersurfaces. In order to solve the constraint (4.6.32), we need expressions for $\overset{3}{R}$ and $E_{ij}$. It is easy to show that

$$\overset{3}{R} = -\frac{4}{a^2}\partial^2\mathcal{R} + \frac{6}{a^2}(\partial\mathcal{R})^2 + \frac{16}{a^2}\mathcal{R}\partial^2\mathcal{R}, \tag{4.6.35}$$

and a short calculation gives

$$E^{ij}E_{ij} - E^2 = -6\left(H + \frac{\dot{\mathcal{R}}}{1+2\mathcal{R}}\right)^2 + 4\left(H + \frac{\dot{\mathcal{R}}}{1+2\mathcal{R}}\right)\overset{3}{\triangle}\psi + \psi^{|ij}\psi_{|ij} - (\overset{3}{\triangle}\psi)^2, \tag{4.6.36}$$

where $\overset{3}{R}$ is the $h_{ij}$ Laplacian. After expanding to first order, we get

$$\mathcal{O}(1): \quad \frac{6H^2}{\kappa^2} - \dot{\phi}^2 - 2V = 0 \tag{4.6.37a}$$

$$\mathcal{O}(\mathcal{R}): \quad -\frac{4}{a^2\kappa^2}\partial^2\mathcal{R} - \frac{12\alpha H^2}{\kappa^2} + \frac{12H}{\kappa^2}\dot{\mathcal{R}} - \frac{4}{a^2\kappa^2}H\partial^2\psi - 2\alpha\dot{\phi}^2 = 0, \tag{4.6.37b}$$

the first of which is the Friedmann constraint, and the second of which can be solved by taking

$$\alpha = \frac{\dot{\mathcal{R}}}{H} \quad \text{and} \quad \psi = -\frac{\mathcal{R}}{H} + \chi, \quad \text{where} \quad \partial^2\chi = \frac{\kappa^2a^2}{2H^2}\dot{\mathcal{R}}\dot{\phi}^2. \tag{4.6.38}$$

Substituting all of this back into the action, we obtain

$$I^{(0)} = \frac{1}{2}\int a^3(1+3\mathcal{R}+\frac{3}{2}\mathcal{R}^2)(1+\frac{\dot{\mathcal{R}}}{H})\mathcal{L}^{(1)} + a^3(1+3\mathcal{R}+\frac{3}{2}\mathcal{R}^2)(1-\frac{\dot{\mathcal{R}}}{H}+\frac{\dot{\mathcal{R}}^2}{H^2})\mathcal{L}^{(-1)}, \tag{4.6.39}$$

where

$$\mathcal{L}^{(1)} = -\frac{4}{a^2\kappa^2}\partial^2\mathcal{R} + \frac{6}{a^2\kappa^2}(\partial\mathcal{R})^2 + \frac{16}{\kappa^2 a^2}\mathcal{R}\partial^2\mathcal{R} - 2V \tag{4.6.40a}$$

$$\mathcal{L}^{(-1)} = -\frac{6}{\kappa^2}\left(H + \frac{\dot{\mathcal{R}}}{1+2\mathcal{R}}\right)^2 + \frac{4}{\kappa^2}\left(H + \frac{\dot{\mathcal{R}}}{1+2\mathcal{R}}\right)\left(\frac{1}{a^2}(1-2\mathcal{R})\partial^2\psi + \frac{1}{a^2}\partial\psi\partial\mathcal{R}\right) + \dot{\phi}^2 \tag{4.6.40b}$$

After expanding to second order, dropping first order terms because they are proportional to the background equations of motion, and collecting the remaining pieces, we have

$$I^{(0)} = \frac{1}{2}\int a^3\left[\frac{2}{\kappa^2 a^2}(\partial\mathcal{R})^2 - \frac{4}{\kappa^2 a^2}\frac{\dot{\mathcal{R}}}{H}\partial^2\mathcal{R} + \mathcal{R}^2\left(-3V - \frac{9H^2}{\kappa^2} + \frac{3\dot{\phi}^2}{2}\right)\right.$$
$$\left. + \frac{\mathcal{R}\dot{\mathcal{R}}}{H}\left(-6V + \frac{6H^2}{\kappa^2} - 3\dot{\phi}^2\right) + \dot{\mathcal{R}}^2\frac{\dot{\phi}^2}{H^2}\right]. \tag{4.6.41}$$

By integrating by parts, it is easy to show that

$$\int -\frac{4}{\kappa^2}\frac{a}{H}\dot{\mathcal{R}}\partial^2\mathcal{R} = \int \frac{2a}{\kappa^2}\left(1 - \frac{\dot{H}}{H^2}\right)\mathcal{R}\partial^2\mathcal{R}, \tag{4.6.42}$$

and after integrating the $\mathcal{R}\dot{\mathcal{R}}$ term by parts and using the background equations of motion to simplify the result, we obtain (4.6.25) without any necessity to invoke the slow roll approximation. We can now recall that $\zeta = -\mathcal{R}$ during scalar field inflation to rewrite the result in terms of $\zeta$.

Eq. (4.6.25) can be cosmetically tidied up by introducing a new variable $u = z\zeta$. In terms of $u$,

$$I^{(0)} = \int \mathrm{d}\tau\, \mathrm{d}^3x\, \frac{1}{2}\left(u'^2 - u_{,i}u_{,i} + \frac{z''}{z}u^2\right). \tag{4.6.43}$$

which reproduces the field equation (4.6.23). Incidentally, this shows that the dynamics of $\zeta$ are really determined fairly directly by the effective field theory for $Q$. Where $Q$ obeys a less trivial effective field theory, exactly the same argument holds and allows a quick derivation of the Mukhanov equation for $\zeta$.[27]

The action for $u$ is equivalent to the theory of a real scalar field propagating on Minkowski space with time-dependent effective mass $m^2_{\text{eff}} = z''/z$. It is a tedious but fairly straightforward calculation to show that $m^2_{\text{eff}}$ can be expressed in terms of $aH = \mathcal{H}$

[27] For comparison, see, eg., the complicated manipulations involved in Garriga and Mukhanov (1999).

and the slow-roll parameters (Lidsey, Liddle, Kolb, Copeland, Barreiro, and Abney, 1997; Stewart and Lyth, 1993),

$$m_{\text{eff}}^2 = 2(aH)^2\left(1+\varepsilon-\frac{3}{2}\eta+\varepsilon^2-2\varepsilon\eta+\frac{1}{2}\eta^2+\frac{1}{2}\xi^2\right), \tag{4.6.44}$$

where $\varepsilon$ was defined in (4.5.15) and $\eta$ and $\xi$ satisfy

$$\eta = \frac{M_{\text{P}}^2}{4\pi}\frac{H''}{H} \quad \text{and} \quad \xi = \frac{M_{\text{P}}^2}{4\pi}\left(\frac{H'H'''}{H^2}\right)^{1/2}. \tag{4.6.45}$$

To fix the values of $\varepsilon$, $\eta$, $\xi$ and $aH$ we will need the details of the classical solution around which this quantum theory is a perturbation. In general, there are no analytical solutions of the Einstein–scalar field system; one must work in terms of approximations and perturbation expansions. However, there is one choice of scalar field potential for which the theory is exactly solvable. By performing a perturbation expansion around this point we can calculate $m_{\text{eff}}^2$ in an open neighbourhood of the exact theory. The idea is to solve (4.5.13) for $H(\phi)$. Make the Ansatz

$$\dot{\phi} = -\sqrt{\frac{M_{\text{P}}^2}{4\pi p}}H, \tag{4.6.46}$$

where $p > 1$ is a characteristic number, and the numerical prefactor and factors of $M_{\text{P}}$ are chosen to make the slow roll parameters come out simply. Substitution in (4.5.13) allows one to integrate immediately for $H(\phi)$, giving

$$H(\phi) = H_0 \exp\left(\sqrt{\frac{16\pi}{pM_{\text{P}}^2}}\phi\right). \tag{4.6.47}$$

One can then integrate the Ansatz to find $\phi$ as a function of $t$, if desired. This solution is known as power law inflation, because the scale factor $a$ is proportional to $t^p$. The slow-roll parameters are

$$\varepsilon = \eta = \xi = \frac{1}{p}. \tag{4.6.48}$$

In an open neighbourhood of this theory, the slow-roll parameters need be neither constant nor equal, but their deviation from each other must be small. If we assume in addition that each slow-roll parameter is individually small, then one can solve for $aH$,

$$\tau = \int \frac{\mathrm{d}t}{a} = -\frac{1}{aH}(1+\varepsilon) + O(\text{slow-roll}^2). \tag{4.6.49}$$

This assumption, that $O(\max(\varepsilon, |\eta|)) \ll 1$, is known as the slow-roll approximation. It is ubiquitous in the theory of inflation, but it is not a necessary condition and is made only

for the purpose of obtaining analytical approximations. (It is easy to see that slow-roll is sufficient for inflation to occur, since $\varepsilon < 1$ is the definition of inflation.) Keeping only terms which are first-order in a slow-roll parameter shows that the effective mass is

$$m_{\text{eff}}^2 = \frac{2 + 6\varepsilon - 3\eta}{\tau^2} = \frac{\mu^2 - 1/4}{\tau^2}. \tag{4.6.50}$$

where $\mu = 3/2 + 2\varepsilon - \eta$ is a constant. The quantum scalar action is therefore

$$I^{(0)} = \frac{1}{2}\int \mathrm{d}\tau\, \mathrm{d}^3x \left[(\partial_\tau u)^2 - \delta^{ij}\partial_i\partial_j u + \frac{\mu^2 - 1/4}{\tau^2}u^2\right] = \frac{1}{2}\int \mathrm{d}\tau\, \mathrm{d}^3x\; u\Box_\mu u, \tag{4.6.51}$$

which is a free, Gaussian theory with kernel

$$u\Box_\mu u = u\left[\overleftarrow{\partial_\tau}\overrightarrow{\partial_\tau} - \delta^{ij}\overleftarrow{\partial_i}\overrightarrow{\partial_j} + \frac{\mu^2 - 1/4}{\tau^2}\right]u. \tag{4.6.52}$$

The other term in the semiclassical action is the graviton contribution, which corresponds to a spin 2 perturbation of the metric, in contrast to the spin 0 terms we have been considering. (In principle there can be vorticity modes corresponding to the spin 1 sector, but these die away rapidly, like $a^{-4}(p+\rho)^{-1}$ (Liddle and Lyth, 1993).) The action is (Mukhanov et al., 1992)[28]

$$I^{(2)} = \frac{1}{8\kappa_4^2}\int \mathrm{d}\tau\, \mathrm{d}^3x\; a^2\partial_c h^{ab}\partial^c h_{ab}, \tag{4.6.53}$$

where $h_{ab}$ is the spin 2 field corresponding to a metric perturbation

$$\mathrm{d}s^2 = a^2(\delta_{ab} + h_{ab})\mathrm{d}x^a\mathrm{d}x^n. \tag{4.6.54}$$

After rescaling the field, so that $m_{ab} = (4\kappa_4^2)^{-1/2}ah_{ab}$, one obtains the action

$$I^{(2)} = \frac{1}{2}\int \mathrm{d}\tau\, \mathrm{d}^3x\; m^{ab}\Box_\nu m_{ab}. \tag{4.6.55}$$

This is a free, Gaussian theory with the same form as the scalar action, but with an effective mass given instead by $m_{\text{eff}}^2 = a''/a$, so

$$m_{\text{eff}}^2 = \frac{\nu^2 - 1/4}{\tau^2} \tag{4.6.56}$$

[28]This action is nowhere near as complicated to obtain as the $I^0$ contribution. One merely expands the Einstein action to second order in $h$, which is tedious but not difficult, integrates by parts, and drops terms which are total derivatives.

with $\nu = 3/2 + \varepsilon$.[29]

For any quantum field $\phi$, the variance $\sigma_\phi^2$ (or dispersion (Liddle and Lyth, 1993)) at a point $x$ in spacetime is defined by the rule[30]

$$\sigma_\phi^2(x) = \lim_{y \to x} \langle \Omega | \mathrm{T}\phi(y)\phi^\dagger(x) | \Omega \rangle = \lim_{y \to x} \Delta_F(x, y), \tag{4.6.58}$$

where T is the time-ordering symbol, and $\Delta_F(x, y)$ is the Feynman propagator, or dressed propagator in the sense of (2.3.27). The power spectrum corresponding to this variance is

$$\Delta_\phi^2(k) = \frac{\mathrm{d}\sigma_\phi^2}{\mathrm{d}\ln k} \tag{4.6.59}$$

which is a function of the wavenumber $k$. A spectral index can also be defined via

$$n_\phi = \frac{\mathrm{d}\ln \Delta_\phi^2(k)}{\mathrm{d}\ln k}. \tag{4.6.60}$$

---

[29]The description of microphysical degrees of freedom carried in inflation by a trivial Gaussian theory of this sort means that the final perturbation spectrum will also be Gaussian, ie., characterized entirely by its two-point function. If one considers more sophisticated models where coupling to other fields are not neglected, then one can obtain non-Gaussian signatures (Gupta, Berera, Heavens, and Matarrese, 2002). However, there is little observational motivation for such a complication of the scenario, because present microwave background measurements favour Gaussianity (Spergel et al., 2003).

[30]We are being naïve about the product of operators at a point. As explained in Chapter 2, the product of operators at a point is usually singular, admitting only a kind of Laurent expansion in the separation $d$ as $d \downarrow 0$. (This is the operator product expansion; see p. 27.) In this case, that means that the propagator $\langle \Omega | \mathrm{T}\phi(y)\phi^\dagger(x) | \Omega \rangle$ may exhibit short-distance singularities as $y$ approaches $x$. The short-distance behaviour is dominated by small wavelengths in the Fourier representation, or large wavenumbers $k$, so it is the $k \to \infty$ behaviour which determines how singular the propagator is in the limit. One might therefore worry about a potential ill-definedness of the power spectrum (4.6.59). On the other hand, it is not exactly the propagator itself which we are interested in, but rather its logarithmic $k$-space derivative. Since differentiation reduces the order of divergence by one (that is, if a quantity diverges like $k^n$ as $k \to \infty$, then its derivative typically diverges like $k^{n-1}$), repeated differentiation generally brings enough factors of $k$ into the integrand to cause it to converge (and this is, in fact, a practical method of regularizing a quantum field theory; see, eg., Weinberg (1994)). In the present case, (4.6.59) only means that the short-distance limit of the propagator can be reconstructed via

$$\sigma_\phi^2 = \int \mathrm{d}\ln k \, \Delta_\phi^2(k) + A \tag{4.6.57}$$

where $A$ is a constant which is typically infinite. There may also be singularities in the other extreme, where the propagator connects two points which are separated by very long distances and the behaviour is dominated by modes of small wavenumber. The presence of masses in the deep infra-red usually acts to regulate such divergences, but if the theory is empty in the infra-red then the propagator will diverge there.

In practice, this means that we approximately fitting the amplitude via the rule

$$\Delta^2_\phi(k) = A^2_{\phi|k_0}\left(\frac{k}{k_0}\right)^{n_\phi} \tag{4.6.61}$$

over a local range of $k$, where $k_0$ is a pivot wavenumber in the régime of interest, $A^2_{\phi|k_0}$ is the amplitude at $k = k_0$, and the approximation is only good over a finite range of $k$.

These definitions are conventional, except for the matter power spectrum (see also (4.3.8)), which was defined as $\mathcal{P}(k) = \langle|\delta_{\mathbf{k}}|^2\rangle$. This is related to $\Delta^2_\delta(k)$ by the rule[31]

$$\Delta^2_\delta(k) = \frac{k^3}{2\pi^2}\mathcal{P}(k). \tag{4.6.62}$$

(This follows from the definition of $\Delta^2_\delta$ as the logarithmic derivative of the variance, because $\sigma^2_\delta = \int \mathrm{d}^3k\,(2\pi)^{-3}\,|\delta_{\mathbf{k}}|^2$, and the integral is isotropic.) In a longitudinal gauge with zero shear, $\zeta$ effectively becomes the gravitational potential $\Phi$ and satisfies Poisson's equation,

$$\triangle\Phi = 4\pi G a^2\rho\delta, \tag{4.6.63}$$

which in Fourier space is equivalent to the rule

$$\zeta_{\mathbf{k}} \sim \frac{(aH)^2}{k^2}\delta_{\mathbf{k}} \quad \text{and so} \quad \Delta^2_\zeta \sim \frac{\mathcal{P}(k)}{k}. \tag{4.6.64}$$

In order to make the spectral index $n_\zeta$ coincide with the index $n$ in the matter power spectrum (4.3.8), one chooses

$$n_\zeta - 1 = \frac{\mathrm{d}\ln\Delta^2_\zeta(k)}{\mathrm{d}\ln k}. \tag{4.6.65}$$

To calculate the power spectrum of the fields $\phi$ or $m_{ab}$, one begins with the Gaussian kernel (4.6.52). To invert this kernel, one transforms to Fourier space,

$$\Box_\mu = \int \frac{\mathrm{d}^3x}{(2\pi)^3}\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\left(-\vec{\partial}^2_\tau - k^2 + \frac{\mu^2 - 1/4}{\tau^2}\right), \tag{4.6.66}$$

so the inverse operator must be

$$\Box^{-1}_\mu = \frac{1}{(2\pi)^3}\int \mathrm{d}^3p\,\mathrm{e}^{-\mathrm{i}\mathbf{p}\cdot\mathbf{x}}G_F(\mathbf{k};\tau^2), \tag{4.6.67}$$

[31] We observe that there are no hard and fast rules about notating power spectra, and many different and mutually incompatible conventions are used simultaneously in the field. Consequently, any recourse to the literature is inevitably confusing. The power spectrum $\Delta^2$ is sometimes denotes $\mathcal{P}^2$ or $\mathcal{P}$, and sometimes just $\Delta$. We reserve $\mathcal{P}$ to describe the matter power spectrum, the $\mathbf{k}$-space amplitude $\langle|\delta_{\mathbf{k}}|^2\rangle$, whereas this quantity is sometimes written $P$. The canonical resource for matter power spectrum issues is the book by Liddle and Lyth (2000) (see also Liddle and Lyth (1993)), who use the $\mathcal{P}$ convention for the power spectrum, and denote square roots explicitly when required; they make no use of the symbol $P$ (our $\mathcal{P}$).

where $G_F$ satisfies

$$\left(-\overrightarrow{\partial}_\tau^2 - k^2 + \frac{\mu^2 - 1/4}{\tau}\right) G_F = \delta(\tau - \eta). \tag{4.6.68}$$

Evidently $G_F$ must be continuous at $\tau = \eta$, so integrating over a small neighbourhood of this point shows that

$$[\partial_\tau G_F]_-^+ = -1, \tag{4.6.69}$$

where $[f(z)]_-^+$ at a point $z = z_0$ is defined by

$$[f(z)]_-^+ = \lim_{z \to z_0^+} f(z) - \lim_{z \to z_0^-} f(z). \tag{4.6.70}$$

Away from $\tau = \eta$, $G_F$ satisfies the homogeneous form of (4.6.68), which has solutions of the form

$$\alpha L_\mu^{(1)}(-k\tau) + \beta L_\mu^{(2)}(-k\tau) \tag{4.6.71}$$

in which $L^{(1,2)}$ satisfies

$$L_\mu^{(1,2)}(z) = z^{1/2} H_\mu^{(1,2)}(z). \tag{4.6.72}$$

The most general solution for $G_F$ can be parametrized according to the rule[32]

$$G_F = \frac{\pi \mathrm{i}}{4k} \begin{cases} aL_\eta^{(1)} L_\tau^{(1)} + bL_\eta^{(2)} L_\tau^{(1)} + cL_\eta^{(1)} L_\tau^{(2)} + dL_\eta^{(2)} L_\tau^{(2)} & (\tau < \eta) \\ eL_\eta^{(1)} L_\tau^{(1)} + fL_\eta^{(2)} L_\tau^{(1)} + gL_\eta^{(1)} L_\tau^{(2)} + hL_\eta^{(2)} L_\tau^{(2)} & (\tau > \eta), \end{cases} \tag{4.6.73}$$

where $a$, $b$, $c$, $d$, $e$, $f$, $g$ and $h$ are arbitrary complex numbers which turn out to parametrize the vacuum (we shall call them 'vacuum parameters') and $L_\tau^{(n)}$ is an abbreviation for $L^{(n)}(-k\tau)$. Continuity at $\tau = \eta$ and the jump condition (4.6.69) require various relations among the eight vacuum parameters. From continuity we learn that

$$b + c - f - g = 0, \quad a + d - e - h = 0, \quad \text{and} \quad a - d - e + h = 0, \tag{4.6.74}$$

whereas from the jump condition one obtains

$$b - c - f + g = 2. \tag{4.6.75}$$

Since this quantum field is perfectly well behaved, the propagator must be a Hermitian operator. Imposing Hermiticity of $G_F$ supplies the additional constraints $a = h^*$, $b = g^*$,

[32]This most general of path integral derivations does not seem to have appeared in the literature before.

$c = f^*$ and $d = e^*$.[33] Using these Hermiticity results to reduce the equations which resulted from continuity yields somewhat simpler restrictions,

$$\mathrm{Im}(c) = -\mathrm{Im}(b), \quad \mathrm{Im}(a) = -\mathrm{Im}(d), \quad \text{and} \quad \mathrm{Re}(a) = \mathrm{Re}(d). \tag{4.6.77}$$

There is also the jump condition, which says $\mathrm{Re}(b) - \mathrm{Re}(c) = 1$. One can parametrize the remaining freedom in a number of ways. A popular parametrization (see Hwang (1995) where this parametrization is introduced, although not in these terms) is to set

$$a = c_1 c_2^*, \quad d = c_1 c_2^*, \quad b = |c_1|^2, \quad \text{and} \quad c = |c_2|^2, \tag{4.6.78}$$

in which case the vacuum condition is

$$|c_1|^2 - |c_2|^2 = 1. \tag{4.6.79}$$

The choice of the complex numbers $c_1$ and $c_2$ determine what constitutes the quantum vacuum at asymptotically late and early times. This is easiest to see in the canonical picture, where $c_1$ and $c_2$ parametrize the elementary wavefunctions – or modes of the field equation – out of which one builds the general quantum field $\phi$. The problem of selecting the correct vacuum state was discussed in Section 4.5.3, where it was pointed out that new physics at the Planck scale could be expected to change our naïve view of the vacuum. The problem appears here in its concrete form: excitations in the fields are measured by the coefficients of the elementary wavefunctions, so different choices of $c_1$ and $c_2$ lead to different measurements of the field excitations (Birrell and Davies, 1982; Wald, 1994) (see later, in Section 5.7.2). In Minkowski space, one can rely on Poincaré invariance as a crutch to shore up one's low energy, Newtonian intuition of particles, but this is not available in general, and quite arbitrary solutions of Einstein's equations do not exhibit any specific symmetry group which could be used to discriminate between competing bases.

[33] To see how this works in detail, observe that $\langle \mathrm{T}\phi(x)\phi^\dagger(y)\rangle = -\mathrm{i}G_F(x,y)$, so taking Hermitian conjugates shows that *opposite* propagator satisfies

$$\langle \mathrm{T}\phi(y)\phi^\dagger(x)\rangle = \mathrm{i}G_F^*(x,y). \tag{4.6.76}$$

This is nothing more than Feynman's famous observation that a particle propagating from $x$ to $y$ is indistinguishable from an antiparticle propagating backwards in time from $y$ to $x$. In this case the time ordering implicit in the Feynman propagator $G_F$ means that no exchange of $x$ and $y$ is necessary, so $G_F(x,y) = -G_F^*(x,y)$, from which the stated result follows.

The dilemma is real, and unlikely to be swept away until a full quantum gravity is available to address such foundational issues.

In the present cosmological case, however, there is a limit in which Poincaré invariance re-emerges and one can appeal to low energy intuition to preserve us from the perils of quantum gravity. In the small scale limit, where the field cannot probe the curvature of space, one might reasonably expect field excitations to appear which obey the familiar dispersion relation $E^2 = \mathbf{p}^2 + m^2$, which is a consequence of Poincaré invariance. This expectation means that the elementary wavefunctions out of which the quantum Hilbert space is built should approach their Minkowski counterparts as $k \to \infty$. Such a choice demands $c_2 = 0$ and corresponds to an adiabatic vacuum in the sense of Birrell and Davies (1982). It is often called the Bunch–Davies vacuum, and can be considered a canonical choice for inflation. As described in Section 4.5.3, some recent work has centred on the possibility of other vacua. This different vacua can be phenomenologically described by mutilating the conventional dispersion relation, but appear naturally in the theory by making different choices of $c_1$ and $c_2$. A popular parametrization scheme in this context is the method of $\alpha$-vacua (Danielsson, 2002b), where the constants $c_1$ and $c_2$ are described by

$$c_1 = \frac{1}{\sqrt{1 - \mathrm{e}^{\alpha+\alpha^*}}} \quad \text{and} \quad c_2 = \frac{\mathrm{e}^{\alpha}}{\sqrt{1 - \mathrm{e}^{\alpha+\alpha^*}}}. \tag{4.6.80}$$

In this $\alpha$-parametrization, the Bunch–Davies vacuum corresponds to $\alpha = -\infty$. These different possible choices of vacuum may imprint signatures on inflationary spectra. Generically, one expects such effects to be order $H/\Lambda$, where $\Lambda$ is the ultra-violet cutoff during inflation (Danielsson, 2002a).[34]

Taking the coincidence limit of the propagator gives the variance,

$$\sigma^2(\tau) = \lim_{\tau \to \eta} \mathrm{i} \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{\pi \mathrm{i}}{4k} \left( c_1 c_2^* L^{(1)} L^{(1)} + |c_1|^2 L^{(1)} L^{(2)} + |c_2|^2 L^{(2)} L^{(1)} + c_1^* c_2 L^{(2)} L^{(2)} \right), \tag{4.6.81}$$

where an extra factor of i has been added to make the Green's function $G_F$ correspond to the full propagator, since this theory is still Lorentzian. For cosmological applications, one is interested in the very large scale limit $|k\tau| \to 0$. In this régime, the variance goes over to

$$\sigma^2(\tau) = -\frac{1}{8\pi^3} \int \frac{k^2 \, \mathrm{d}k}{k} \, |c_1 - c_2|^2 \left[ 2^{\mu} \Gamma(\mu) (-k\tau)^{1/2-\mu} \right]^2 \tag{4.6.82}$$

[34] See also the discussion of vacuum parametrization and squeezing in Appendix C.

- The matter power spectrum. Eq. (4.6.25) shows that $\zeta = -u/z$, so choosing the Bunch–Davies vacuum, one has

$$\sigma_\zeta^2 = \frac{1}{z^2}\frac{1}{8\pi^3}\int \frac{k^2\,\mathrm{d}k}{k}\,\left[2^\mu\Gamma(\mu)(-k\tau)^{1/2-\mu}\right]^2. \tag{4.6.83}$$

After substituting for $z$ and evaluating on the horizon scale $k = aH$, we obtain the familiar result (Birrell and Davies, 1982; Liddle and Lyth, 2000; Peacock, 1999)

$$\Delta_\zeta^2 = \left[\left.\frac{H^2}{|H'|}\right|_{k=aH}\frac{2^{\mu-1/2}}{M_\mathrm{P}^2}\frac{\Gamma(\mu)}{\Gamma(3/2)}(\mu-1/2)^{1/2-\mu}\right]^2. \tag{4.6.84}$$

At zero order $\mu = 3/2$, which corresponds to taking the background spacetime to be exactly de Sitter. In this special case, one has

$$\Delta_\zeta^2 = \frac{M_\mathrm{P}^4}{16\pi^2}\frac{H^4}{H'^2}. \tag{4.6.85}$$

The spectral index $n_\zeta$ is easily evaluated,

$$n_\zeta - 1 = \left.\frac{\mathrm{d}\ln\Delta_\zeta^2(k)}{\mathrm{d}\ln k}\right|_{k=aH} \simeq \frac{\mathrm{d}\ln\Delta_\zeta^2(k)}{\mathrm{d}\ln a} = -4\varepsilon + 2\eta. \tag{4.6.86}$$

Since $\varepsilon$ and $\eta$ are small, this predicts a primordial power spectrum close to the scale invariant Harrison–Zel'dovich spectrum discussed above.

- The graviton power spectrum. For each polarization separately, the gravitational perturbation $h$ is related to $u$ by

$$h = \sqrt{\frac{32\pi}{M_\mathrm{P}^2}}\frac{u}{a}, \tag{4.6.87}$$

so the variance in $h$ is given by

$$\sigma_h^2 = \frac{32\pi}{M_\mathrm{P}^2}\frac{1}{a^2}\frac{1}{8\pi^3}\int \frac{k^2\,\mathrm{d}k}{k}\left[2^\nu\Gamma(\nu)(-k\tau)^{1/2} - \nu\right]^2. \tag{4.6.88}$$

Evaluating on the horizon scale, truncating the result to zero order as above, and summing over both polarization modes gives the expected answer,

$$\Delta_g^2 = \frac{16H^2}{\pi M_\mathrm{P}^2}. \tag{4.6.89}$$

The spectral index is

$$n_g = -2\varepsilon. \tag{4.6.90}$$

The next step is to relate the power spectrum we have just calculated, which is a quantum-mechanical expectation value, with the power spectrum we measure in the universe today, which is a property of a classical Gaussian random field.[35] This is not at all a trivial operation. The general theory of the quantum-to-classical transition is outlined in Appendix C.

## 4.7. The no-hair theorem

Before moving on to consider current observational data, this is a convenient place to summarize the theory of cosmological observables which has been presented so far. The theory is based on spatially homogeneous, isotropic solutions of Einstein's equations. When coupled to suitable input from particle physics (particle content, interactions cross sections, and masses) and thermodynamics, this leads to predictions about the thermal behaviour of gross matter in the universe. Small perturbations in the matter mix lead to density variations which grow at rates depending on the ambient background and eventually condense into the delicate web of galaxies, clusters and superclusters we see in the universe around us. These perturbations also left an imprint on the relativistic radiation background – the cosmic microwave background – and give specific predictions for temperature anisotropies on the CMB sky.

The observed flatness of the universe and the isotropy of the microwave background are difficult to explain in this standard picture. For this purpose, one introduces an epoch of evolution during which the density parameter $\Omega$ is driven to unity. During this epoch, perturbations are laid down in the matter and gravitational wave sectors, which seed the primordial power spectrum of matter fluctuations. Inflation can also be invoked to dilute away any population of dangerous cosmological relics, such as topological defects arising from any phase transition in the particle physics at high energies which breaks whatever gauge group is effective there to the Standard Model. The general paradigm seems successful, but one would like some reassurance that while the universe is being driven to spatial flatness, any initial irregularities are washed away and the universe is brought to an observationally pristine state of homogeneity and isotropy in preparation for the close-to-scale-invariant inflationary perturbation spectrum to be laid down.

[35] We have not much discussed the statistics of the power spectrum, but they are assumed to be Gaussian in the large.

A rigorous result in this direction was given in an early paper by Wald (1983) and still remains essentially the strongest statement which can be made. In order to make definite statements, it is necessary to make some assumptions about the matter theory, less any degrees of freedom which are only contributing to the effective energy–momentum tensor. One assumes both the strong and weak energy conditions,

$$\begin{aligned} T_{ab}n^a n^b &\geqslant 0 \quad \text{(weak energy condition)} \\ \left(T_{ab} - \frac{1}{2}g_{ab}T\right) n^a n^b &\geqslant 0 \quad \text{(strong energy condition)} \end{aligned} \tag{4.7.1}$$

Let $\Sigma$ by a spatial hypersurface with unit future-pointing normal $n^a$. The connexion $\nabla$ on spacetime induces a preferred connexion $\hat{\nabla}$ on $\Sigma$, out of which one can build a curvature via the Ricci rule,

$$[\hat{\nabla}_a, \hat{\nabla}_b]Y_c = \overset{3}{R}_{abcd}Y^d \tag{4.7.2}$$

for all vectors $Y_d$. The curvature $\overset{3}{R}$ is related to the spacetime curvature $R$ via the Gauss–Codacci equations,

$$\begin{aligned} \overset{3}{R}_{ab} &= h_a{}^c h_b{}^d T_{cd} + h_a{}^c h_b{}^d n^e n^f R_{ecfd} - KK_{ab} + K_{ac}K_b{}^c \\ \overset{3}{R} &= R + 2R_{ab}n^a n^b - K^2 + K^{ab}K_{ab} \end{aligned} \tag{4.7.3}$$

where $K_{ab}$ is the extrinsic curvature,

$$K_{ab} = h_a{}^c h_b{}^d \nabla_c n_d \tag{4.7.4}$$

and $h_{ab}$ is a projection tensor onto $\Sigma$, $h_{ab} = g_{ab} + n_a n_b$. Let us agree to decompose $K$ into a dilation $\theta$ and a traceless tensor $\sigma$ describing the shear and vorticity,

$$K_{ab} = \sigma_{ab} + \frac{1}{3}\theta h_{ab} \tag{4.7.5}$$

Using (4.7.3) and the Einstein equation gives a constraint equation,

$$\theta^2 - 3\Lambda = -\frac{3}{2}\overset{3}{R} + \frac{24\pi}{M_{\mathrm{P}}^2}T_{ab}n^a n^b + \frac{3}{2}\sigma_{ab}\sigma^{ab}. \tag{4.7.6}$$

One can also find an evolution equation for the dilation $\theta$. Since $\theta = \nabla^a n_a$, the derivative $\dot{\theta} = n^a \nabla_a \theta$ satisfies

$$\dot{\theta} = n^b \nabla_b \nabla_a n^a = n^b \nabla^a \nabla_b n_a - R_{ab}n^a n^b = \nabla^a \dot{n}_a - (\nabla_a n_b)(\nabla^a n^b) - R_{ab}n^a n^b. \tag{4.7.7}$$

Substituting the decomposition of $K_{ab}$ gives the Raychaudhuri equation

$$\dot{\theta} = \nabla^a \dot{n}_a - \sigma^{ab}\sigma_{ab} - \frac{1}{3}\theta^2 - R_{ab}n^a n^b. \tag{4.7.8}$$

Using the Einstein equation we can again rewrite this in terms of the energy–momentum tensor

$$\dot{\theta} - \Lambda + \frac{1}{3}\theta^2 = \nabla^a \dot{n}_a - \sigma^{ab}\sigma_{ab} - \frac{8\pi}{M_{\rm P}^2}\left(T_{ab} - \frac{1}{2}T g_{ab}\right) n^a n^b. \tag{4.7.9}$$

It is a geometrical fact that $\sigma^{ab}\sigma_{ab} \geqslant 0$. In all Bianchi spacetimes except Bianchi type IX, the curvature of spatial sections satisfies $\overset{3}{R} < 0$, so applying the weak energy condition to the constraint equation and the strong energy condition to the Raychaudhuri equation gives

$$\dot{\theta} \leqslant \Lambda - \frac{1}{3}\theta^2 \leqslant 0. \tag{4.7.10}$$

Evidently $\theta \geqslant \sqrt{3\Lambda}$ for all time. Using this and the time integral of the first inequality shows that

$$\sqrt{3\Lambda} \leqslant \theta \leqslant \frac{(3\Lambda)^{1/2}}{\tanh(t\sqrt{\Lambda/3})}. \tag{4.7.11}$$

Taking the limit as $t \to \infty$ shows that $\theta \to \sqrt{3\Lambda}$ at long times. Subsituting for $\theta$ in the constraint equation shows that

$$\sigma_{ab}\sigma^{ab} \leqslant \frac{2}{3}(\theta^2 - 3\Lambda) \to 0 \quad \text{as } t \to \infty, \tag{4.7.12}$$

so $\sigma_{ab} \to 0$ as $t \to \infty$.

The no-hair theorem shows that provided there is an effective cosmological constant and the matter theory obeys the energy conditions, then any initial perturbations decay, the three-geometry approaches spatial flatness, and the spatial sections become isotropic on a timescale $(3/\Lambda)^{1/2}$. (Since the vorticity perturbations are not sourced, this justifies out neglect of spin 1 modes during inflation.) One expects the mechanism for this decay to resemble that for the black hole Israel theorem: as inflation proceeds, inhomogeneities are carried over the de Sitter horizon and disappear, and the ball of observable spacetime surrounding any observer settles down to an accurately de Sitter space.

## 4.8. Dark energy and the anthropic landscape

Modern cosmology is built on the foregoing account of thermal history, large scale structure and an inflationary epoch, with the addition of an extra novel feature. The traditional inflationary picture of cosmology envisions a vacuum energy dominated epoch in the very early universe, which subsequently decays and reheats the universe with a mixture of matter and radiation that survives to the present day. In this picture, the late universe quietly cools as the gravitational redshift continuously carries energy out of the

radiation. However, recent observations give support to the idea that the late universe may also contain a very small vacuum component. This component would have been subdominant from the end of inflation until roughly the present day and does not interfere with the accounts given above, which largely ignored this possibility. The first hints that a sizable vacuum components was present appeared from supernova Hubble diagrams which suggested a deviation from the naïve Hubble law at large redshifts, and these results have now been conclusively confirmed by a slew of experimental data, including cosmic microwave background tests, galaxy clustering and large-scale structure, and cluster x-ray emission. Many of these sources are independent, and the results are in remarkable agreement, so there seems little room for doubt that vacuum energy really is present in nature.

The possibility of a late-time cosmological constant is not a new idea. The Einstein equations themselves can be mutilated by adding to the Einstein tensor a term proportional to $\Lambda_{CC} g_{ab}$ which obeys the Bianchi condition and cannot be excluded merely on grounds of symmetry or simplicity. In this case $\Lambda_{CC}$ would be an entirely new constant of nature that may eventually be explained by appeal to string theory or whatever theory of high energy physics controls excitations in the ultra-violet, and would survive from the earliest ages until the present day. If $\Lambda_{CC}$ is small, then this might explain the small cosmological constant that observations apparently detect. There is nothing really wrong with this idea, except that it does not really explain anything, since the mysterious constant $\Lambda_{CC}$ with dimensions of (energy)$^4$ has no microphysical origin, and also there is the suggestion of deeper physics at work because the transition to an epoch of the universe's evolution in which we could actually observe $\Lambda_{CC}$ has occurred only comparatively recently. In the very distant past there would have been no possibility of seeing a small cosmological constant, since it would have been swamped by the energetic sea of matter and radiation propagating over it, and in the distant future all matter and radiation will have been diluted away and only the cosmological constant would remain. The present time is a very special epoch at which both matter and radiation *and* $\Lambda_{CC}$ are visible, and the coincidence begs explanation.

As an alternative, we can seek anthropic enlightenment. The anthropic principle has appeared in cosmology before, with varying success, but unfortunately our lack of understanding of quantum gravity, and in particular the initial conditions of the universe, means

that when attempting to apply the principle we are driven to arguments which can seem to have little to do with physics. There are strong opinions on both sides, to the point where it sometimes appears arguments are evaluated on their philosophical content rather than their scientific merit. Nevertheless, there are some conclusions which can be reached with comparative confidence. At the outset, it is clear that vacuum domination can occur only relatively late in any universe which is capable of supporting life sophisticated enough to enquire about its secrets. This happens because the onset of $\Lambda$-domination essentially kills off the growth of bound structures, so if the cosmological constant becomes dominant too early, then large scale structures such as galaxies and galaxy groups would never form, and the carbon-based life they support in the local neighbourhood would never have appeared. Therefore if the cosmological constant is not zero, and it seems that it is not, it cannot be very large. This is the weakest form of the anthropic principle: the universe is the way it is, because we grew up to be able to find out about it.

It is possible to be more quantitative, although only at the expense of some generality and an increase in model-dependence (Sahni and Starobinsky, 2000). An early anthropic argument for $\Lambda \neq 0$ was given by Banks (1985) and Weinberg (1987). This argument, given by Weinberg in its most sophisticated form, shows that large values of $\Lambda$ are unlikely to be observed if the presence of observers demands the existence of galaxies. A more elaborate variation on the same theme is the following (Martel, Shapiro, and Weinberg, 1998). Suppose that we have an ensemble of observers living in a given region of the universe (or a given sub-universe if we wish to suppose that such things exist), where the vacuum in that region (or sub-universe) selects a particular cosmological constant, number of visible space-time dimensions, amount of CP violation, number of flavours, and the symmetry group of gauge interactions, among other things. The probability that these observers will measure a value of $\rho_\Lambda$ for the vacuum energy is postulated to be

$$\mathbb{P}(\rho_\Lambda) = \frac{F(\rho_\Lambda)}{\int_0^\infty \mathrm{d}\rho_\Lambda \, F(\rho_\Lambda)}, \tag{4.8.1}$$

where $F(\rho_\Lambda)$ is the fraction of matter in galaxies in the region of the universe with vacuum energy $\rho_\Lambda$. (The value of $F(\rho_\Lambda)$ can be calculated by assuming Gaussian initial fluctuations at recombination, and normalizing the spectrum to the WMAP results.) If we then require that the observed value of $\Omega_\Lambda^*$ in our own region of space equals the *statistical* mean or

median evaluated over all possible regions, so that $\Omega_\Lambda^* = \langle \Omega_\Lambda \rangle$, then this peaks in the region $\Omega_\Lambda^* \sim 0.6 - 0.9$ for a broad region of parameter space.

Of course, these anthropic arguments aside, it is entirely possible that there is nothing here more profound than a remarkable coincidence in the same way that the Moon and the Sun coincidentally have the same projected size as viewed from the Earth. In the past this was not the case, and eventually it will not be the case in the future, but for the present this coincidence allows the spectacular possibility of solar eclipses. But if we adopt the idea that such a remarkable fact cannot be a simple coincidence, then the idea of just mutilating the Einstein equations to obtain the right value of $\Lambda_{CC}$ will not do, and we will have to seek alternative explanations.

One rather attractive alternative is to detach ourselves from the arduous business of building microphysical models and seek explanations in abstraction and universality. One proposal along these lines is the idea of holographic inflation, which describes the evolution of the universe as a renormalization group flow between two fixed points which describe conformal field theories. The dS/CFT correspondence, if it exists, translates such CFTs into de Sitter gravitational states. This proposal is outlined in Appendix D. A similar but more speculative proposal is to use an ad hoc inter-brane potential in the late universe to generate a small positive vacuum energy (see the discussion of the cyclic scenario in Section 5.3). On the other hand, there is the obvious strategy that one can apply exactly the same technology used in inflation – that is to say, a light or massless weakly coupled scalar field whose vacuum energy dominates the energy density of the universe – to supply the necessary effective cosmological constant. This has the advantage that the relevant physics is fairly well understood and comparatively robust, although it is not without its problems.

Distinguishing between the various competing proposals is far from trivial (Melchiorri and Odman, 2002). In such a case, one can resort to mere phenomenological parametrization in the hope that the information obtained in this way will eventually inform our ideas about microphysics. The most important parametrization of the vacuum energy is its

equation of state[36]

$$p_\Lambda = w\rho_\Lambda, \tag{4.8.2}$$

where $w$ is an observationally measurable number. Notation and nomenclature in this area is not yet standardized, and there are a selection of competing usages. In particular, it is common to hear the vacuum energy described as dark energy, although the name is poor since the vacuum energy is no less luminous or more energetic than any other form of potential energy. Similarly the late-time scalar field which plays the role of the inflaton in the present-day universe is often called quintessence, even though the "fifth force" which such a massless field might mediate would be no different to the situation during inflation, and no standard matter particle apparently carries its charge.

The parameter $w$ must be less than $-1/3$ in order that the universe inflate, and greater than or equal to $-1$ in order to preserve causality (Hawking and Ellis, 1973). The cosmological constant itself has $w = -1$. This does not prevent models in which $w < -1$ from being considered; such material is sometimes called phantom matter. We do not consider phantom matter in this thesis. Present observations constrain $w$ to be rather close to $-1$ (Bean and Melchiorri, 2002; Melchiorri, Mersini, Odman, and Trodden, 2003), and there is not yet any guarantee that the equation of state is in fact any different from that of a pure cosmological constant, although there is some hope that a discrimination might be accessible in the not too distant future.[37] A further discriminant is possible evolution in $w$, which does not occur for a genuine cosmological constant. Any signature of evolution is a tell-tale sign that one is dealing not with a new constant of nature such as $\Lambda_{CC}$ but instead some microphysical process which gives a similar effect.

We will consider quintessence in somewhat more detail in Chapter 6 and describe some constraints, as well as deriving new ones. There are other comparable models such as so-called k-essence in which the necessary vacuum energy is supplied by the kinetic energy of a scalar field rather than its potential energy. These models can be tuned to help the coincidence problem described above become rather more understandable, although it is rather less than clear whether the problem can be solved entirely by this method.

[36]This is a departure from previously accepted usage, where the equation of state is usually written $p = \omega\rho$. However the substitution of $w$ for $\omega$ is now so firmly established that it is most sensible to count $w$ as one of the standard cosmological parameters.

[37]Current constraints on $w$ are described in the next section, which outlines the current observational position.

If one does not adopt one of the field theory models for dark energy, then the observed smallness of the cosmological constant becomes a significant difficulty. It is not so much that producing a small cosmological constant that is troublesome, but stabilizing the value against significant radiative corrections. Loops in the matter theory will generically contribute vacuum energies that are of order $\Lambda_{\rm UV}^4$, where $\Lambda_{\rm UV}$ is the ultra-violet cutoff of the theory. For matter theories coupled to general relativity this will be of order $M_{\rm P}^4$ or $M_{\rm SUSY}^4$ if one believes that supersymmetry will introduce new physics above the SUSY scale. In theories with fermions one can employ the extra minus signs introduced by fermion loops to cancel some fermionic contributions against bosonic ones, but the cancellation will not be perfect except in cases of exact supersymmetry, leaving residual vacuum energies which are still of order $\Lambda_{\rm UV}^4$. Since supersymmetry is clearly not preserved in the low-energy world, this is not a viable route to understanding the origin of a small cosmological constant in the present day universe.

All of the mechanisms we have discussed so far rely on a dynamical configuration which drives the cosmological constant to its current value. The difficulty with all of these methods roughly amounts to the problematical business of arranging the final effective value of the $\Lambda_{CC}$ to lie in the observationally acceptable range. On the other hand, a well-known alternative strategy is to construct an ensemble of universes. If the effective value of $\Lambda_{CC}$ can take a wide variety of values over the members of this ensemble, then a weak anthropic argument can be invoked to help understand the small value that we actually see. One can argue a great deal about the merits of this line of argument, and many people have over the years, but apart from personal taste and the prevailing theoretical prejudice there is no reason of principle why this sort of construction should not actually exist, and we see a number of examples even in the local universe (for example, the positions of planets around stars that could conceiveably support life). Despite these remarks, it is rather unclear whether this can be a complete answer. For example, unless the alternative vacua are observationally accessible to us, it is not obvious what we have gained by proposing the existence of alternative worlds in which our own existence would not be viable but which would have no immediate experimental consequences.

Clearly any vindication of this scenario must rely on indirect evidence and deduction. However, this general scheme receives some support from modern ideas in string theory, in which the different vacua of M-theory, understood as a global covering theory for all the

individual perturbative superstring theory vacua, are connected. By varying the moduli fields which describe these vacua one can move around in the so-called M-plane, or vacuum manifold of M-theory. This gives a concrete realization of the ensemble of vacua. However, in itself this does not help explain the situation with the cosmological constant, since on the connected sheet of vacua supersymmetry is exactly preserved and the cosmological constant is set to zero. On the other hand, there will generically be a family of disconnected vacua in which supersymmetry is not preserved. This kind of scenario has recently been proposed as the basis for an anthropic explanation of the smallness of the cosmological constant (Susskind, 2003, 2004), building on work by Bousso and Polchinski (2000).

In this proposal, one considers vacua which contain a large number of branes. These branes act as sources for the fields in the theory, of which we assume there are $N$, and their expectation values are quantized according to a generalized Dirac law (Bousso and Polchinski, 2000). These expectation values combine with a bare cosmological constant $\Lambda_0$ to produce a total effective value

$$\Lambda = \Lambda_0 + \frac{1}{2}\sum_{i=1}^{N} n_i^2 q_i^2, \tag{4.8.3}$$

with $n_i$ the excitation level of the $i$'th flux and $q_i$ the corresponding charge gap. With a sufficiently large number $N$ of fluxes, there are sufficiently large number of possible states that it is statistically likely we can obtain a $\Lambda$ in the required observable range without fine tuning. No special conditions are required, but only a large number of possible ways to make the energy. Unfortunately, it is difficult to be certain that the various approximations made in constructing this model are safe because supersymmetry is not available. In this case one relies on the anthropic principle to select the vacuum with small $\Lambda$ as the world in which we live.

This is not the only mechanism available in string theory to construct a vacuum with a small cosmological constant. For example, Kane, Perry, and Zytkow (2003) argue that the true vacuum state will consist of a superposition which mixes the available M-theory vacua, for which no appeal to anthropic arguments is necessary.

## 4.9. Observational summary

Having documented the collection of basic observational quantities that it is the goal of physical cosmology to measure, it is appropriate to give a short summary of the current

observational position. The number of experiments planned or actually underway is very large, so we will focus only on recent, large-scale projects and suggest where improvements may be made from experiments due to return science data in the near future. It is no exaggeration to say the experiments in question – the 2dF galaxy redshift survey (2dF-GRS), the Sloan digital sky survey (SDSS) and the Wilkinson Microwave Anisotropy Probe (WMAP) – represent probably the most significant advance in quantitative cosmology since the discovery of the CMB (Penzias and Wilson, 1965). In addition, the Supernova Cosmology Project earlier furnished us with the first tantalizing observational hints that the expansion of the universe might (against expectation) be accelerating. This wealth of new data has changed the way cosmology as a mature science is conducted. Whereas in earlier times most detailed modelling necessarily had to be founded in plausibility arguments and imprecise or fuzzy data, one may now reasonably expect to know important cosmological observables to a few significant figures, and, more importantly, quote meaningful error bars, although problems with degeneracies still exist. Advances in survey design and simulation techniques allow experimentalists to understand statistical and instrumental errors, and greater computing power coupled with increasingly refined CMB and fluid dynamics codes give highly sophisticated models and predictions with which to compare the outcome of experiments.

Among the supporting cast of planned or future experiments, it is important to mention the gravitational wave observatories LIGO and GEO. These facilities are not designed to measure gravity waves in any part of parameter space which is expected to be relevant to cosmology (instead, they are designed to measure point sources of strong waves such as colliding black holes or black hole–neutron star interactions rather than the stochastic, Gaussian background predicted by inflation), but the mere confirmation that gravity waves exist will open an important new window on cosmological data. Other important upcoming experiments include QUEST (Bowden et al., 2004), the $Q$ and $U$ extra-galactic sub-millimetre telescope, which will attempt to pin down with some precision the polarization of the CMB recently detected by DASI and WMAP and currently beginning science observations at the South Pole, and the European Space Agency satellite experiment Planck which will determine other CMB observables to a satisfying precision. QUEST and Planck constitute our best hope of seeing cosmological signatures of gravity waves in the near future. Like WMAP, Planck data is likely to require combination with other

datasets to provide the best constraints. This is a quite general feature of observation in cosmology. A final major experiment which should appear within the decade is the Large Hadron Collider at CERN, a particle collider which replaces the older lepton-based LEPII experiment. This instrument operates at centre-of-mass energies around 14 TeV, although because of the large QCD backgrounds involved in colliders based on flavour physics the reliable energy window is likely to be limited to a TeV or so. For its intended purpose of discovering supersymmetry this is presumably amply sufficient, since SUSY or some other physics must intervene at energies below a TeV in order to preserve unitarity of certain Higgs processes.

The importance of the LHC for cosmology may not be as apparent as the pure cosmological-observable experiments such as QUEST and Planck, but its consequences for inflation could be profound. On the one hand, the LHC is likely to tell us whether low energy supersymmetry exists, and, if so, could provide information about inflationary models based on generic supersymmetry features such as $F$-term and $D$-term inflation (not discussed in this thesis owing to lack of space; see, eg., Liddle and Lyth (2000); Lyth and Riotto (1999)). If the LHC rules out supersymmetry but instead points to different physics operating at about a TeV – such physics, even if not SUSY, must exist – then this may provide different cosmological clues. More optimistically, it is possible as an outside chance that the LHC may see low scale stringy physics associated with warped compactifications such as those to be discussed in subsequent chapters. This might happen, for example, if the effective scale of gravity is sufficiently low that we begin to see five-dimensional processes occurring, such as the leaking of energy into large extra dimensions, or the formation of microscopic black holes either as end-products or as intermediate states in QCD processes.

The present observational climate presents a strongly optimistic picture. Data is arriving at a faster rate than ever before, and there are solid grounds for optimism that the cumulative picture will provide the necessary tools to allow us to begin to crack the remaining difficulties with the Standard Cosmological Model.

**4.9.1. Galaxy surveys.** The numbers that WMAP would return were known fairly well in advance of the data, in part due to partial-sky balloon-borne or ground-based CMB experiments (such as Boomerang and Maxima) prior to WMAP, but also because of galaxy clustering experiments such as the 2-degree field galaxy redshift survey. Although this survey has been superseded as the largest galaxy survey by the Sloan digital sky survey,

| | |
|---|---|
| Hubble parameter $h$ | $h = 0.665 \pm 0.047$ |
| matter density $\Omega_m$ | $\Omega_m = 0.313 \pm 0.055$ |
| CDM density $\Omega_{\mathrm{CDM}}$ | $\Omega_{\mathrm{CDM}} h^2 = 0.115 \pm 0.009$ |
| baryon density $\Omega_{\mathrm{B}}$ | $\Omega_{\mathrm{B}} h^2 = 0.022 \pm 0.002$ |
| tensor-to-scalar ratio $r$ | $r < 0.7$ at 95% confidence |
| equation of state of vacuum $\omega$ | $\omega < -0.52$ at 95% confidence |

**Table 2.** Cosmological parameters as measured by the 2dF survey

mature data from 2dF was originally used to compare with and supplement WMAP data, so we discuss this redshift survey first.

The 2dF instrument is a robotic instrument sited at the Anglo-Australian Observatory which measured the redshifts of approximately 220,000 galaxies between 1995 and 2002. The experimental data was generally released in June 2003, and early results for cosmological parameters were presented in Percival et al. (2001). Cosmological experiments are almost universally afflicted with degeneracies, in the sense that observations are unable to probe the cosmological parameters individually, but only specific combinations. Thus, for example, an experiment may be unable to probe the baryon density $\Omega_{\mathrm{B}}$ and the Hubble parameter $H = 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ individually, but only the degenerate combination $\Omega_{\mathrm{B}} h^2$. When combined with other data, such as Boomerang and Maxima, these degeneracies can be broken, leading to predictions for the parameters of the cosmological model (Efstathiou et al., 2001). We present details from the full data set (Percival et al., 2002) in Table 2.

**4.9.2. Cosmic microwave background.** The general situation regarding cosmic microwave background experiments is summarized in Bucher, Moodley, and Turok (2000, 2002).

The Wilkinson Microwave Anisotropy Probe is a NASA satellite launched in June 2001. It observes from the Lagrange point L2 which allows the satellite to observe the full sky from a stable orbit with minimal station-keeping manoeuvers. WMAP has spent two years at L2, and (at the time of writing) the second-year science data is presently awaited. The first-year results are summarized in Table 3, and the CMB anisotropy map – presently the best picture of the CMB anisotropy – is reproduced in Figure 4.1.

| | **WMAP** | **WMAP + 2dFGRS + Lyman $\alpha$** |
|---|---|---|
| Hubble parameter $h$ | $h = 0.72 \pm 0.05$ | $h = 0.72 \pm 0.03$ |
| matter density $\Omega_m$ | $\Omega_m h^2 = 0.14 \pm 0.02$ | $\Omega_m h^2 = 0.133 \pm 0.006$ |
| baryon density $\Omega_\mathrm{B}$ | $\Omega_\mathrm{B} h^2 = 0.024 \pm 0.001$ | $\Omega_\mathrm{B} h^2 = 0.0226 \pm 0.0008$ |
| matter spectral index $n_s$ | $n_s = 0.93 \pm 0.03$ | |

**Table 3.** Cosmological parameters as measured by the 2dF survey

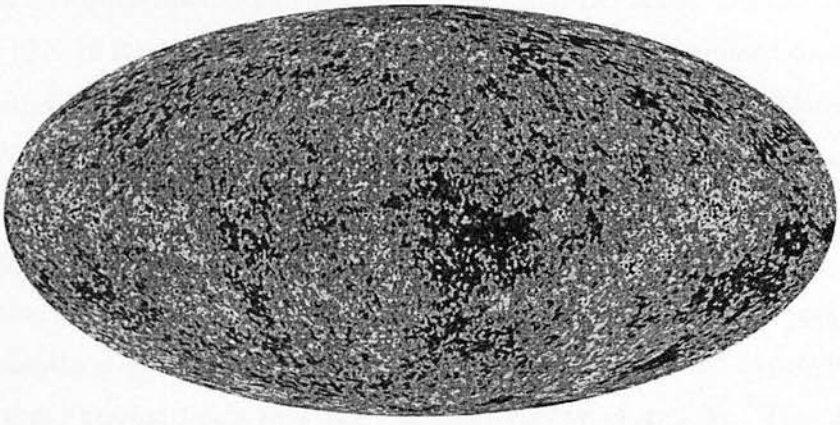


**Figure 4.1.** All-sky map of the cosmic microwave background anisotropy made by the WMAP satellite (picture retrieved from `http://map.gsfc.nasa.gov/m_or.html` on 19 August 2004)

CHAPTER 5

# Brane cosmology

The cosmology described in Chapter 4, and on which all current cosmological estimates, predictions, and observations are based, has remained virtually unchanged (except in the matter inventory) since the early days of Friedmann and Lemaître. On the one hand, this is a definite virtue of the theoretical framework within which the standard cosmology is set: once one assumes homogeneity and isotropy, as are implied by the observations, one is fixed with an FRW universe, and the gross thermal history and perturbative arguments follow just on the basis of what we currently understand about particle physics in laboratories and accelerators here on earth.

On the other hand, general relativity provides no guidance about what sort of manifolds (potentially solutions of the Einstein equations) can really be expected to provide consistent background gravitational fields in a full quantum theory of gravity. The most obvious manifestation of this failure is the prediction of singularities by the Einstein equations (Hawking and Ellis, 1973), or points in spacetime where the worldlines of observers moving in the spacetime may come to an end. Such predictions are presumably not physical and do not correspond so much to a definite singular prediction as a fuzzy prediction of high-energy, high-curvature regions where the quantum dynamics of quantum gravity are important.

One can try to sanitize the situation by asking about the predictions of the Einstein equations subject to various reasonable hypotheses about the matter content (Hawking and Ellis, 1973; Visser, 1996). However, one finds that no matter how stringent one makes the restrictions there are always circumstances in which one expects singularities to appear. When applied to the univese at large, such theorems should not be taken to imply that a cosmolgical singularity event, a 'creation' of the universe, necessarily occurs,[1] but rather than the standard picture of a universe evolving from a early hot dense phase is fairly

[1]In any event, the universe is now known to contain a particular sort of matter (vacuum energy) which violates some of the assumptions under which these theorems were originally proved. This matter would have been cosmologically less important in the past, and the same argument holds: the singularity theorems may no longer hold rigorously, but they were always statements about classical general relativity and not

generic. Before the hot dense phase, little can be said with confidence. These difficulties are not restricted to cosmological solutions in general relativity. For these reasons, one does not know what kind of solutions one is looking for from the Einstein equations. Although the universe we observe does appear four-dimensional, homogeneous and isotropic, one does not have any idea, from the Einstein equations alone, whether the real background gravitational field of the universe ought to include deviations from these properties at some particular energy scale.

Evidently, general relativity must be supplemented with some extra physics in order to provide mechanisms which discriminate between sensible gravitational fields and unphysical fields which are only an artefact of the breakdown of Einstein gravity in the far ultra-violet. This new physics is associated with an ultra-violet completion of the theory which is grafted onto the low energy field theory and defines quantum gravitational dynamics at high energies, and could take the form of new fields or matter content or excitations, or superselection rules forbidding unwanted transitions, or new symmetry groups, or new geometrical effects. Various possibilities have been suggested over time, but one by one almost all proposals have been eliminated, leaving only two plausible candidates: string theory, and loop quantum gravity. In this thesis, the focus will be on extra physics arising from the use of string theory as the ultra-violet completion of gravity plus the standard model.

There are several possible approaches. One can attempt to build a theory of cosmology directly, in string theory. This method is highly technical and it is extremely difficult to carry any calculations to the point of giving definite predictions for the kind of cosmological observables discussed in Chapter 4. However, much progress has recently been made, including the long-standing problem of constructing a de Sitter state within string theory (Kachru et al., 2003b). This opens up the possibility of studying inflation while carrying along the full set of string excitations, although the difficulties in prosecuting this programme are not to be underestimated, and progress is likely to be slow. A more profitable approach has been to begin with stringy backgrounds containing D-branes, and use this to model the high energy régime where inflation is supposed to occur. (There is presumably not much benefit following the stringy dynamics to low energies, where we expect that

---

the real physical world. As such, it was never the case that they unambiguously predicted a singularity in the past of the real universe.

general relativity plus standard standard particle physics will be a good approximation.) Here, the idea is not so much to study cosmology as a dynamical *solution* of string theory as to investigate facets of the extra physics string theory provides, applied to cosmological situations (Dvali and Tye, 1999; Garcia-Bellido, 2003; Garcia-Bellido et al., 2002). One can construct some quite elaborate inflationary scenarios along these lines. A third approach is to begin with a known, consistent string background and truncate it consistently to low energies, usually by throwing away all information about massive modes, to obtain an infra-red effective field theory (Lidsey, Wands, and Copeland, 2000; Lukas, Ovrut, Stelle, and Waldram, 1999a,b; Lukas et al., 1999c). Because almost all the known backgrounds in string theory for which this approach is feasible are supersymmetric, it is difficult to study dynamical problems, since the background is usually a BPS state.

A final option is to combine the best of the last two approaches. One works with stringy physics which can be applied profitably to cosmology, but working within the context of low energy theory, where sophisticated calculational tools are available, and the passage to genuine observables can be effected without excessive difficulty. This allows one to begin the search for new signatures of high energy physics in the observational data, even while the theoretical effort of bringing formal calculations involving such physics to the point where they can make concrete predictions is still underway. Over the next several years, it seems highly likely that the major driver of progress in cosmology will come from observation and not theoretical progress, so this approach is far from speculative: instead, it is pragmatic, economical and may yield large rewards. Rather like our current understanding of inflation, these rewards may eventually need to be understood in the context of an embedding within fundamental physics, but there is no reason why solid progress may not be made purely in terms of phenomenology. This is the principal objective of this thesis.

In this chapter the aim is to describe in detail some brane world scenarios. This serves to provide a sufficient background to place the models that will be developed in Part 2 in context, both within the relatively specialised area of brane compactifications in general, and also within the larger arena of phenomenological models of the early universe. We calculate a number of quantities, such as geometrical invariants of the background spacetimes, which will be useful throughout the course of this thesis, and to which we will frequently refer. We also review the construction of Kaluza–Klein theories on the

brane world, which ties together the various aspects of quantum field theory and string compactification which have previously been discussed.

### 5.1. Binétruy–Deffayet–Ellwanger–Langlois models

The prototypical brane cosmology is the weak truncation of the Hořava–Witten model discussed in the previous chapter, in which the $E_8$ super-Yang Mills matter is replaced by an arbitrary cosmological fluid and eleven dimensional supergravity in the bulk space is replaced by low energy Einstein gravity. This truncation can occur at several levels of rigour. The truncation of the full Hořava–Witten scenario to four dimensions was described comparatively early, beginning with the full eleven dimensional supergravity, and keeping only zero modes after projecting down to four dimensions, Lukas et al. (1999a,b,c). We will describe this model in Section 5.2 below. Scenarios with large extra dimensions were suggested by Arkani-Hamed, Dimopoulos, and Dvali (1998) and Antoniadis, Arkani-Hamed, Dimopoulos, and Dvali (1998) (see also Arkani-Hamed, Dimopoulos, Dvali, and Kaloper (2000)), and by other authors. Much more interest followed the suggestion of Randall and Sundrum (1999a) that large extra dimensions might naturally explain the large hierarchy between the Planck or GUT scale and the electroweak theory. Later, the scenario was refined to suggest an alternative to compactification (Randall and Sundrum, 1999b), in which the extra unwanted dimensions imposed by string theory might not have to be wrapped up on a compactification manifold, as described in Section 3.4. A different weak truncation, with rather different objectives and motivations is the Ekpytotic or cyclic model, which will be the subject of Section 5.3. In this section we describe a class of metric, Einstein gravity braneworlds which are often described as Randall–Sundrum models but which in reality are somewhat more general than the original proposal of Randall and Sundrum (1999b). We shall call them Binétruy–Deffayet–Ellwanger–Langlois models (BDEL models) since the general theory of metric braneworlds of this type was first laid bown by Binetruy, Deffayet, Ellwanger, and Langlois (2000a) (see also Binetruy, Deffayet, and Langlois (2000b)).

Despite the interest it generated, the Randall–Sundrum scenario does not bear much relation to string theory. It is a model constructed by solving the Einstein equations, appropriately coupled to matter, and as such belongs to the phenomenological approach rather than any rigorous (and ambitious) programme to derive cosmology from string

theory. However difficult it might eventually prove to connect the low-energy Randall–Sundrum world with ideas about high energy physics, the suggestion sparked a large-scale effort to study the phenomenology of simple brane worlds based, as Randall & Sundrum had done, only on low energy physics. These are the BDEL models. Exact inflating BDEL solutions were supplied by Kaloper (1999) (see also Kaloper and Linde (1999)), and later a class of exact solutions for an arbitrary cosmological history $H(t)$ (Binetruy et al., 2000a,b). It is these exact solutions which form the background spacetimes for physics to be considered in Part 2, and which will be described in Section 5.1.1 below.

Having obtained a set of exact solutions with which to work, there are two major routes through which progress is likely. One can either press ahead with the speculative theoretical effort of using physics in the new geometry to explain mysterious or peculiar phenomena we see in the low energy world (such as the apparent vacuum energy, or three generations of fermions), which is the route in which we shall be most interested, or one can attempt to carry over the kind of perturbation theory which was described in Section 4.3–4.4 and Section 4.6.1. There is already a formidable literature concerned with the attempt to build an observational perturbation theory in the braneworld, much of which is summarised in convenient form in a recent review (Maartens, 2004). We do not attempt a detailed description of perturbation theory in the braneworld, which is somewhat tangential to our main concern.

**5.1.1. The background BDEL metric.** Consider the almost-factorizable metric (Binetruy et al., 2000a,b)

$$\mathrm{d}s^2 = -n^2(t,y)\mathrm{d}t^2 + a^2(t,y)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j + \mathrm{d}y^2 \tag{5.1.1}$$

In models where the bulk is empty, this spacetime solves Einstein's equations with a cosmological constant, but more generally there may be source terms on and off the brane,

$$R_{ab} - \frac{1}{2}Rg_{ab} = G_{ab} = \Lambda g_{ab} - \delta_{\mathrm{D}}(y)\lambda + \kappa_5^2[Q_{ab} + S_{ab}\delta_{\mathrm{D}}(y)]. \tag{5.1.2}$$

Here, $\Lambda$ is the bulk cosmological constant ($\Lambda > 0$ for de Sitter-like solutions and $\Lambda$ ¡ 0 for anti-de Sitter like solutions), whereas $\lambda$ is the tension on the brane. $Q_{ab}$ is a bulk source and $S_{ab}$ is the energy–momentum tensor of whatever matter and gauge degrees of freedom are stuck to the brane. If $Q_{ab}$ is of compact support, then this is an asymptotically anti-de Sitter (AdS) space, and if $Q_{ab} = 0$ then it is exactly Schwarzschild anti-de Sitter

space, possibly in the degenerate limit where the Schwarzschild mass is zero. The brane is considered to be embedded at $y = 0$.

Solving Einstein's equations (5.1.2) produces solutions for the metric fields $a$ and $n$. One can either calculate the Einstein tensor $G_{ab}$ explicitly, or choose a more elegant approach in which one works with the Einstein–Hilbert action instead. We will employ solutions based on the action principle in Chapter 8, and since the historical development began with the Einstein tensor, we present that approach here. The governing equations for the fields $a$ and $n$ are the $(0,0)$, $(5,5)$ and $(5,0)$ Einstein equations,

$$\frac{a''}{a} + \left(\frac{a'}{a}\right) - \frac{1}{n^2}\left(\frac{\dot{a}}{a}\right)^2 = -\frac{\Lambda}{3}\delta^0_0 + \frac{\lambda}{3}\delta^0_0\delta_{\rm D}(y) + \frac{\kappa_5^2}{3}\left[Q^0{}_0 + S^0{}_0\delta_{\rm D}(y)\right] \tag{5.1.3a}$$

$$\left(\frac{a'}{a}\right)^2 - \frac{1}{n^2}\left(\frac{\dot{a}}{a}\right)^2 + \frac{n'}{n}\frac{a'}{a} + \frac{1}{n^2}\frac{\dot{n}}{n}\frac{\dot{a}}{a} - \frac{1}{n^2}\frac{\ddot{a}}{a} = -\frac{\Lambda}{3}\delta^5_5 - \frac{\kappa_5^2}{3}Q^5{}_5 \tag{5.1.3b}$$

$$\frac{n'}{n}\frac{\dot{a}}{a} - \frac{\dot{a}'}{a} = -\frac{\Lambda}{3}\delta^5_0 - \frac{\kappa_5^2}{3}Q^5{}_0. \tag{5.1.3c}$$

Integrating the first of these over a small neighbourhood of $y = 0$ produces the jump condition for $a$, which can also be obtained using the Israel condition (Israel, 1966),

$$\frac{[a']^+_-}{a_{\rm b}} = \frac{\lambda}{3} - \frac{\kappa_5^2}{3}\rho, \tag{5.1.4}$$

where we have written the brane source $S^a{}_b$ as a perfect fluid, $S^a{}_b = \mathrm{diag}(-\rho, p, p, p)$, which is the appropriate choice for a cosmology, and quantities evaluated on the brane are written with a subscript b, thus, for example $a_{\rm b} = a(y = 0)$. For reasonable choices of the bulk source $Q^a{}_b$ the right-hand side of (5.1.3c) is zero, so $n$ is determined to within an arbitrary function of $t$,

$$n(t, y) = \frac{\dot{a}(t, y)}{\alpha(t)}. \tag{5.1.5}$$

Eq. (5.1.5) can be understood as a manifestation of the reparametriztion invariance of the theory (Chapter 8). If one assumes that the bulk sources $Q_{ab}$ decouple, either by setting them to zero or by fixing $Q_{ab}$ to be transverse and traceless, then the Einstein equations become

$$\frac{a''}{a} + \left(\frac{a'}{a}\right)^2 - \frac{\alpha^2}{a^2} = -\frac{\Lambda}{3} + \frac{\lambda}{3}\delta_{\rm D}(y) - \frac{\kappa_5^2}{3}\rho\delta_{\rm D}(y) \tag{5.1.6a}$$

$$\left(\frac{a'}{a}\right)^2 - \frac{\alpha^2}{a^2} + \frac{a'}{a}\frac{\dot{a}'}{\dot{a}} - \frac{\alpha}{a}\frac{\dot{\alpha}}{\dot{a}} = -\frac{\Lambda}{3}. \tag{5.1.6b}$$

Incidentally, the fact that $Q_{ab}$ decouples if it is transverse and traceless shows that gravitational waves do not couple to the gross degrees of freedom on the brane.

It will be very convenient to have on hand the brane Friedmann equation, analogous to (4.2.4), which can be obtained without explicitly solving for $a$ and $n$. This apparent peculiarity is explained because, as in the four-dimensional case, the Friedmann equation is really a constraint and not an evolution equation. By rewriting the first equation in (5.1.6a) as an exact differential,

$$\frac{1}{2}\mathrm{d}_y(aa')^2 = \frac{\alpha^2}{2}\mathrm{d}_y a^2 - \frac{1}{4}\frac{\Lambda}{3}\mathrm{d}_y a^4, \tag{5.1.7}$$

one can obtain an immediate first integral,

$$\frac{\alpha^2}{a^2} = \left(\frac{a'}{a}\right)^2 + \frac{1}{2}\frac{\Lambda}{3} + \frac{\mathcal{C}}{a_{\mathrm{b}}^4}, \tag{5.1.8}$$

where $\mathcal{C}$ is an arbitrary constant of integration, which turns out to be related to a possible Schwarzschild-like mass in the bulk. It behaves like a new, non-interacting matter component with a radiative equation of state and is often referred to (for this reason) as dark radiation. For the present, however, we are entirely free to make the choice $\alpha = \dot{a}_{\mathrm{b}}$, in which case evaluating (5.1.8) around the brane[2] and replacing $a'/a$ with its value, as computed via the jump condition (5.1.4), yields the brane Friedmann equation,

$$H^2 = \frac{\Lambda_4}{3} + \frac{\kappa_4^2}{3}\rho\left(1 + \frac{\rho}{2\lambda}\right) + \frac{\mathcal{C}}{a_{\mathrm{b}}^4}. \tag{5.1.9}$$

The four-dimensional cosmological constant $\Lambda_4$ and the four-dimensional gravitational coupling appearing here are defined by the rules

$$\Lambda_4 = \frac{1}{2}\left(\frac{\lambda^2}{6} + \Lambda\right) \quad \text{and} \quad \kappa_4^2 = -\frac{\kappa_5^2}{6}\lambda. \tag{5.1.10}$$

Two useful quantities for future reference are the ratio of the four- and five-dimensional gravitational couplings,

$$\mu = \frac{\kappa_4^2}{\kappa_5^2} = -\frac{\lambda}{6}, \tag{5.1.11}$$

and the Anti-de Sitter curvature scale,

$$\ell = \sqrt{-\frac{6}{\Lambda}} = \frac{1}{\sqrt{\mu^2 - \Lambda_4/3}}. \tag{5.1.12}$$

[2]We ignored the distributional terms when carrying out the first integral, so this is really defined by a limiting procedure.

If the four-dimensional cosmological constant is tuned to vanish, so that $\Lambda_4 = 0$, the the curvature scale is $\ell = \mu^{-1}$. With these choices, the on-brane Friedmann equation can be rewritten,

$$H^2 + \frac{1}{\ell^2} = \mu^2 \left(1 + \frac{\rho}{\lambda}\right)^2 + \frac{\mathcal{C}}{a_{\rm b}^4}, \tag{5.1.13}$$

which is an alternative form that is occasionally useful.

**5.1.2. The metric fields $a$ and $n$.** We have now assembled sufficient machinery to solve for the fields $n$ and $a$ directly. The most convenient route is to introduce a new variable $v = a^2$, in terms of which (5.1.6a) can be written

$$v'' - 2\dot{a}_{\rm b}^2 = \frac{4}{\ell^2} v. \tag{5.1.14}$$

This has the general solution

$$v = A(t) \cosh \frac{2y}{\ell} + B(t) \sinh \frac{2y}{\ell} - \frac{\ell^2}{2} \dot{a}_{\rm b}^2, \tag{5.1.15}$$

where $A$ and $B$ are arbitrary functions of $t$. Let us restrict attention to $y > 0$. This is an exponential-looking profile in $y$, which can have cosh- or sinh-like behaviour, depending on the relative disposition of $A(t)$ and $B(t)$. For example, if this were to behave like $\cosh 2y\ell^{-1}$, then $v$ would be regular and non-zero everywhere for $|y| < \infty$, but the resulting metric would be very peculiar. In particular, proper distances would expand indefinitely near large $|y|$ and the entire volume of the spacetime would be concentrated at infinity. This is unphysical, and can be avoided by displacing the exponential profile using the sinh term. As a result, we are looking for a profile which starts at some value $v_{\rm b}$ on the brane and decreases smoothly to zero at some finite distance in the transverse dimension. This cuts off the divergence of $v$ as $|y| \to \infty$ and keeps a finite volume concentrated near the brane at $y = 0$.

With this in find, notice that to fix $A(t)$ and $B(t)$ it is only necessary to specify a cosmological evolution $H(t)$ on the brane. The difference between (5.1.14) and the other Einstein equation (cf. (5.1.6a)) in the bulk (neglecting distributional terms with support only at $y = 0$) requires that a much simpler equation hold, viz.

$$v'' + \dot{v} - v'\dot{v}' + 4v\dot{a}_{\rm b}\ddot{a}_{\rm b} = 0. \tag{5.1.16}$$

Inserting the solution (5.1.15) yields a restriction on $A$ and $B$,

$$A^2 - B^2 = \frac{\ell^4 \dot{a}_{\rm b}^4}{4} + Z \tag{5.1.17}$$

where $Z$ is an arbitrary constant of integration, arising from the elimination of time derivatives in (5.1.16), which must be related to the constant $\mathcal{C}$ which appeared in the Friedmann equation. Moreover, $A$ can be found directly by specializing (5.1.15) to $y = 0$ and imposing the consistency requirement $v = a_{\rm b}^2$ there,

$$A = a_{\rm b}^2 + \frac{\ell^2 \dot{a}_{\rm b}^2}{2}. \tag{5.1.18}$$

The other unknown function $B$ can be fixed using (5.1.17) and (5.1.13), which yields

$$B^2 = \ell^2 a_{\rm b}^4 \mu^2 \left(1 + \frac{\rho}{\lambda}\right)^2 + \ell^2 \mathcal{C} - Z, \tag{5.1.19}$$

after using the Friedmann equation. Choosing $Z$ to balance the $\ell^2\mathcal{C}$ term[3] and taking square roots leaves

$$B = \pm \ell a_{\rm b}^2 \mu \left(1 + \frac{\rho}{\lambda}\right), \tag{5.1.20}$$

where, as described above, the minus sign must be chosen to keep $a$ bounded as $y \to \infty$. As a result, the general solution for $a$ is

$$a^2 = \left(a_{\rm b}^2 + \frac{\ell^2 \dot{a}_{\rm b}^2}{2}\right) \cosh \frac{2y}{\ell} - \ell a_{\rm b}^2 \mu \left(1 + \frac{\rho}{\lambda}\right) \sinh \frac{2y}{\ell} - \frac{\ell^2 \dot{a}_{\rm b}^2}{2}, \tag{5.1.21}$$

from which $n$ can be found by direct differentiation.

- The Randall–Sundrum model. This is the special case where the cosmology is time independent and $\rho = 0$. Therefore the brane must carry Minkowski space, in which case the four-dimensional cosmological constant must be tuned to vanish. The bulk solution is

$$a^2 = a_{\rm b}^2 \left(\cosh \frac{2y}{\ell} - \ell\mu \sinh \frac{2y}{\ell}\right), \tag{5.1.22}$$

but since $\Lambda_4 = 0$ and therefore $\ell\mu = 1$, this can be much simplified, viz.,

$$a^2 = \text{constant} \times \mathrm{e}^{-2|y|\ell^{-1}}, \tag{5.1.23}$$

where the overall constant scale is unimportant, and just comes from choosing a constant value for $a_{\rm b}$. The metric function $n$ equals $a$. (This requires care if proceeding via (5.1.5).)

[3] If desired, one can be a little more rigorous by using the jump condition to relate $v$ on $y > 0$ to $v$ on $y < 0$ and impose a $\mathbf{Z}_2$ symmetry to evaluate the constants appearing there, or by using the jump condition to explicitly evaluate $Z$. The result is the same.

- Kaloper–Linde inflating model. Alternatively, one can look an analogue of the de Sitter state, where there is a four-dimensional cosmological constant driving an eternal inflationary epoch. The Hubble rate $H$ is fixed, so $a_{\rm b} = {\rm e}^{Ht}$ follows just from the definition of $H$. Moreover, the bulk solution is

$$a^2 = a_{\rm b}^2 \left[ \left( 1 + \frac{H^2 \ell^2}{2} \right) \cosh \frac{2y}{\ell} - \mu \ell \sinh \frac{2y}{\ell} - \frac{\ell^2 H^2}{2} \right]. \tag{5.1.24}$$

The hyperbolic terms can be combined into a single function,

$$a^2 = a_{\rm b}^2 \frac{H^2 \ell^2}{2} \left[ \cosh \frac{2}{\ell}(y - y_{\rm h}) - 1 \right] = a_{\rm b}^2 \frac{H^2 \ell^2}{2} \cosh^2 \ell^{-1}(y - y_{\rm h}), \tag{5.1.25}$$

where $y_{\rm h}$ is a constant defined via

$$\tanh \frac{2y_{\rm h}}{\ell} = \frac{2\mu\ell}{2 + H^2 \ell^2}. \tag{5.1.26}$$

In the literature, the definition $a = a_{\rm b}\mathcal{A}$ is often used, in which

$$\mathcal{A}(y) = \frac{H\ell}{\sqrt{2}} \cosh \ell^{-1}(y - y_{\rm h}). \tag{5.1.27}$$

The metric functions approach zero as $y \to y_{\rm h}$. This location constitutes a coordinate horizon, at which the local Gaussian normal coordinates which were employed in (5.1.1) break down (Bowcock, Charmousis, and Gregory, 2000; Mukohyama, Shiromizu, and Maeda, 2000). (We will re-interpret this horizon in terms of an ultra-violet gauge theory cutoff in the next section.)

There is no anti-de Sitter analogue of the no-hair theorem (Section 4.7). If such a theorem existed, then one might be tempted to speculate that any solution of the Einstein equations with a negative cosmological constant eventually settles down to anti-de Sitter space, and any brane solution would therefore be attracted to the inflating Kaloper–Linde model in the same way that de Sitter space is an attractor state for any four-dimensional cosmology with a non-zero positive cosmological constant (Weinberg, 1972). However, this is not true. This is our first indication that there may be problems with the stability of braneworlds, an idea which we shall return to later.

These models are supposed to be motivated by string theory, and are related to the stringy D-branes which were discussed in the previous chapter. In this context, their construction is fairly natural. D-branes couple only to open strings, which carry gauge theories like the Standard Model. Therefore matter exists only on the branes. On the other hand, the branes do not couple to closed strings at all, so gravitational modes propagate

unrestrictedly in the bulk. In the next sections we study models with a rather more concrete M-theory motivation.

## 5.2. Heterotic M-theory and moving brane models

A step up in sophistication is provided by the proper truncation of the Hořava–Witten model (Lukas et al., 1999a,b,c) to five dimensions, according to the scheme

$$E_8 \times E_8 \text{ string} \overset{\text{strong coupling}}{\longrightarrow} \text{SUGRA on } M_{10} \times \mathbf{S1/Z}_2 \overset{\text{compactify}}{\longrightarrow} M_4 \times K_6 \times \mathbf{S1/Z}_2.$$

The Hořava–Witten model involves supergravity on a particular eleven dimensional background. The compactification, where six dimensions are wrapped up on the Calabi–Yau three-fold[4] $K_6$, will produce Kaluza–Klein fields on the $K_6$ and $\mathbf{S}^1/\mathbf{Z}_2$. Most of these modes will be heavy, but, for some topological classes of manifold there will be zero modes. This

[4]When discussing string compactifications in Section 3.2 we were not very specific about what restrictions on the compactification manifold $K_6$ would lead to interesting four-dimensional physics. Partly this was so that the main points of the discussion would apply equally to field theory, which is going to be the case of principal interest in Part 2, but also because the details are not really important for any of the formalism we are going to develop. However in this case we are being explicit about the topological class of the manifold since the truncation under discussion is supposed to lead to a viable model of our universe. The condition is that the four-dimensional physics should admit $\mathcal{N} = 1$ supersymmetry (Freund, 1988; Galperin et al., 2001; Weinberg, 1994) in the low-energy compactified world.

Supersymmetry does not often appear in this thesis, so there is little need to enter into tedious detail. Supersymmetries are generated by covariantly conserved spinors $\zeta$, satisfying $\nabla_a \zeta = 0$. This is a strong restriction on spacetime, because the spinor Ricci identity implies

$$[\nabla_a, \nabla_b]\zeta = \frac{1}{8} R_{abcd}\gamma^{cd}\zeta, \tag{5.2.1}$$

and if $\nabla_a \zeta = 0$ then $R_{abcd}\gamma^{cd}\zeta = 0$. This can be satisfied only for a very special class of curvature tensor $R_{abcd}$. It is almost (but not quite) the condition that spacetime be flat. The details are technical, but the result can be summarised quite simply: $R_{abcd}\gamma^{cd}\zeta = 0$ is the statement that on parallel transport around a closed loop, $\zeta$ comes back to itself not with an arbitrary rotation in the group $SO(6)$ (the tangent bundle group of $K_6$) but with a rotation in the subgroup $SU(3)$. A manifold with this property is said to have $SU(3)$ holonomy; a manifold with holonomy group $G \subset SO(6)$, with $G$ a proper subset, is said to be of special holonomy. Manifolds of $SU(3)$ holonomy can be constructed from complex manifolds of dimension $d = 2n$. Form complex coordinates $z^i$, $\bar{z}^{\bar{j}}$ and define a metric $G$ which satisfies

$$G_{i\bar{j}} = \frac{\partial}{\partial z^i}\frac{\partial}{\partial \bar{z}^{\bar{j}}} K(z, \bar{z}) \tag{5.2.2}$$

for some function $K$ called the Kähler potential. This metric has $G_{ij} = G_{\bar{i}\bar{j}} = 0$; such a metric is called Hermitian.

happens for bosons (antisymmetric tensor fields or form fields on the manifold) under certain cohomological conditions outlined in Chapter 3. For fermions there are analogous conditions involving the Dirac index, index $\not{\partial}$ (de Azcárraga and Izquierdo, 1995; Green et al., 1987; Greene, 1997). Provided $K_6$ and the $\mathbf{S}^1/\mathbf{Z}_2$ are sufficiently small, the non-zero modes can be safely expected to be very heavy and can consistently be set to zero, although one must be careful about modes on the $\mathbf{S}^1/\mathbf{Z}_2$ (Lukas et al., 1999b). This is the central idea of a consistent truncation: one sets to zero massive modes which are not sourced when the fields are on-shell.

One may then enquire whether the resulting model has any cosmological solutions (Lukas et al., 1999c). The natural home for both the truncation and cosmological solutions is the gauged five-dimensional supergravity constructed by Ceresole and Dall'Agata (2000). This gauged supergravity is an extremely complicated theory whose details are not strongly relevant to the rest of the present thesis. Instead of entering into a long discussion here, we describe the construction of heterotic M-theory models in summary form on a purely ad-hoc basis (indeed, the full five dimensional gauged supergravity had not been constructed when Lukas et al. (1999b,c) appeared), and refer the reader to the literature for more detail.

Before proceeding, it is worth noting that the truncated five-dimensional theory contains a non-zero potential, which arises from a non-zero flux of the four-form field strength of supergravity on the internal Calabi–Yau dimensions $K_6$. The presence of this potential means that, except in the trivial case where the Calabi–Yau decompactifies and the

A manifold has $SU(3)$ holonomy if and only if it is Ricci-flat and Kähler. There is an alternative characterization. The Kähler form $J_{1,\bar{1}}$ is defined by

$$J_{1,\bar{1}} = \mathrm{i}G_{i\bar{j}}\mathrm{d}z^i\mathrm{d}\bar{z}^{\bar{j}}. \tag{5.2.3}$$

One can show that this form is closed, so it identifies a member of the complex de Rham cohomology $H^{1,\bar{1}}$, called the Kähler class of the manifold. Also, the mixed components of the Ricci form

$$R_{1,\bar{1}} = R_{i\bar{j}}\mathrm{d}z^i\mathrm{d}\bar{z}^{\bar{j}} \tag{5.2.4}$$

define an equivalence class in $H^{1,\bar{1}}$. With the normalization $c_1 = R_{1,\bar{1}}/2\pi$ one calls $c_1$ the first Chern class of the manifold (de Azcárraga and Izquierdo, 1995). If the manifold is Ricci-flat then the first Chern class is trivial. Yau's theorem states that any Kähler manifold with vanishing first Chern class admits a unique Ricci-flat metric with a given complex structure and Kähler class (Greene, 1997; Polchinski, 1998). (This had earlier been conjectured by the French mathematician Calabi, and for this reason such manifolds are known as Calabi–Yau manifolds.)

four-form flux disappears, flat space is no longer the vacuum solution of the theory. In four-dimensional general relativity, there is a celebrated and very non-trivial result that the vacuum is stable: this was first proved by technical means by Schoen and Yau, and later a simplified spinorial proof was offered by Witten, which can be understood as the limit of a supegravity argument (Witten, 1981).[5] After further reducing the Hořava–Witten vacuum to four dimensions one can attempt to generalise the spinorial stability proof to the braneworld, but the result is no longer manifestly positive definite. In other words, the stability of Minkowski space does not trivially extend to the braneworld. This can be considered a reflection of the tendency of the vacuum to spontaneously decompactify in the absence of an external stabilizing mechanism. The stability of braneworlds is a very general problem, but it is hardly more serious than the entire question of the stability of string compactifications in general, and a solution of one almost certainly entails a solution of the other. In this thesis no attempt is made to address the stability problem, although we shall return to it from time to time when its consequences obtrude upon our notice. In particular, the instability might be interpreted (at least in the open string sector) in terms of a tachyon field theory (Frolov and Kofman, 2004) in the sense of Sen.

**5.2.1. The effective five-dimensional action.** Following Lukas et al. (1999b), we restrict to bosonic fields and write the action in the form $S = S_{\mathrm{SG}} + S_{\mathrm{YM}}$ where $S_{\mathrm{SG}}$ is the 11-dimensional supergravity action (Freund, 1988),

$$S_{\mathrm{SG}} = -\frac{1}{2\kappa^2}\int_{M^{11}}\sqrt{-g}\left[R + \frac{1}{24}G_{IJKL}G^{IJKL} + \frac{\sqrt{2}}{1728}\varepsilon_{I_1\cdots I_{11}}C_{I_1I_2I_3}G_{I_4\cdots I_7}G_{I_8\cdots I_{11}}\right], \tag{5.2.5}$$

where $C_{IJK}$ is a three-form and $G = \mathrm{d}C$ is its field strength. In addition, if this theory is to describe the low-energy Hořava-Witten world then it should contain the two $E_8$ Yang–Mills theories. These are supported on two $M^{10}$s, as described by $S_{\mathrm{YM}}$ (see also Moss (2003)),

$$S_{\mathrm{YM}} = \sum_{i\in\{1,2\}} -\frac{1}{8\pi\kappa^2}\left(\frac{\kappa}{4\pi}\right)^{2/3}\int_{M^{10}_{(i)}}\sqrt{-g}\left\{\operatorname{Tr}F_{(i)}^2 - \frac{1}{2}\operatorname{Tr}R^2\right\}. \tag{5.2.6}$$

Since $G_{IJKL}$ is a field strength, it should obey the Bianchi identity $\mathrm{d}G = 0$ (see Eq. (B.5.7) et seq.), but this is spoiled because of the topological non-triviality of the background.

[5] In supergravity the Hamiltonian is the modulus-squared of a fermionic operator, and as such is trivially positive-definite.

Although $\mathrm{d}G$ is zero almost everywhere, it has $\delta$-function contributions at the branes,

$$(\mathrm{d}G)_{11IJKL} = -\frac{1}{2\sqrt{2}\pi}\left(\frac{\kappa}{4\pi}\right)^{2/3}\left\{J^{(1)}\delta_{\mathrm{D}}(x^{11}) + J^{(2)}(x^{11}-\pi\rho)\right\}_{IJKL}, \tag{5.2.7}$$

where $\rho$ is the radius of the eleventh dimension. We have already seen a similar effect in the BDEL metric models discussed above. The sources $J^{(i)}$ which sit at the branes are given by

$$J^{(i)} = \mathrm{Tr}\, F^{(i)} \wedge F^{(i)} - \frac{1}{2}\,\mathrm{Tr}\, R \wedge R. \tag{5.2.8}$$

This arrangement is required in order that the theory be supersymmetric and anomaly free (Horava and Witten, 1996a,b; Moss, 2003, 2004).[6] However, the presence of the sources complicates any attempt to find solutions of the equations of motion leading to a low-energy compactification scenario.

To work around this difficulty, there are several possible approaches. On the one hand, there is always perturbation theory. One begins with the explicit metric on $K_6 \times M_4 \times \mathbf{S}^1/\mathbf{Z}_2$,

$$\mathrm{d}s_{11}^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + R_0^2 \mathrm{d}x_{11}^2 + V_0^{1/2}\Omega_{AB}\mathrm{d}x^A\mathrm{d}x^B, \tag{5.2.9}$$

where $\Omega_{AB}$ is a Kähler metric, $\mu$, $\nu$ are indices on four-dimensional Minkowski space, and $A$, $B$ are indices on the Calabi–Yau. One then constructs a solution perturbatively in the coupling $\kappa$. To order $\kappa^{2/3}$, on obtains

$$\mathrm{d}s_{11}^2 = (1+\hat{b})\eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu + R_0^2(1+\hat{\gamma})\mathrm{d}x_{11}^2 + V_0^{1/3}(\Omega_{AB}+h_{AB})\mathrm{d}x^A\mathrm{d}x^B, \tag{5.2.10}$$

[6] Although we are not describing the theory at a level of detail where this is obvious, Hořava and Witten originally constructed their model for the $E_8 \times E_8$ heterotic string by considering anomaly cancellation. Propagating theories of *interacting* spin 3/2 particles usually suffer from anomalies, and only the special conditions imposed by supersymmetry protect the gravitino from quantum inconsistencies (Freund, 1988; Göckeler and Schücker, 1987; Weinberg, 1994). This is not such an arduous business as might be supposed. The details of supergravity on ten dimensional backgrounds are well known: this is just the Green–Schwarz superstring, so the fact that $E_8 \times E_8$ supermatter is sufficient to cancel the gravitational anomaly has been known since the first superstring revolution.

where the fields $\hat{b}$, $\hat{\gamma}$ and $h_{AB}$ depend on the $x^{11}$ and Calabi–Yau coordinates, but not the coordinates on the $M_4$. Explicitly (Lukas et al., 1999b),

$$\hat{b} = -\frac{\sqrt{2}}{3} R_0 V_0^{-2/3} \alpha(|x^{11}| - \pi\rho/2) \tag{5.2.11}$$

$$\hat{\gamma} = \frac{2\sqrt{2}}{3} R_0 V_0^{-2/3} \alpha(|x^{11}| - \pi\rho/2) \tag{5.2.12}$$

$$h_{AB} = \frac{\sqrt{2}}{3} R_0 V_0^{-2/3} \alpha(|x^{11}| - \pi\rho/2) \Omega_{AB}. \tag{5.2.13}$$

In writing these expressions, we have set to zero all heavy modes on the Calabi–Yau which correspond to higher harmonics, keeping only a single massless mode that represents the 'breathing' of the three-fold. In addition, the constant $\alpha$ is

$$\alpha = -\frac{1}{8\sqrt{2}\pi} \left(\frac{\kappa}{4\pi}\right)^{3/2} \left(\int_{K_6} \sqrt{\Omega}\right)^{-1} \int_{K_6} \omega \wedge \mathrm{Tr}\, R^{(\Omega)} \wedge R^{(\Omega)}. \tag{5.2.14}$$

The correction terms scale linearly with distance along the orbifold.

On the other hand, one can attempt to reduce the eleven-dimensional theory to five dimensions first, only then attempting to find the braneworld solution. It turns out that this approach allows one to work non-perturbatively in $\kappa$, subject to the condition that one is dealing only with zero-modes, of course. In this case, one would adopt the metric

$$\mathrm{d}s_{11}^2 = V^{-2/3} g_{\alpha\beta} \mathrm{d}x^\alpha \mathrm{d}x^\beta + V^{1/3} \Omega_{AB} \mathrm{d}x^A \mathrm{d}x^B, \tag{5.2.15}$$

where now $\alpha$, $\beta$ are indices on a five-dimensional manifold. The corrections $\hat{b}$, $\hat{\gamma}$ and $h_{AB}$ seen above are absorbed into the fields $V$ and $g_{\alpha\beta}$. To proceed, one splits the three-form $C_{IJK}$ in familiar Kalauza–Klein fashion according to its lower-dimensional field content, which consists of a five-dimensional three-form $C_{\alpha\beta\gamma}$, with field strength $G_{\alpha\beta\gamma\delta}$, a vector $A_\alpha$ with field strength $F_{\alpha\beta}$, a scalar $\xi$, and a harmonic form $\omega_{ABC}$ on the Calabi–Yau,

$$C_{\alpha AB} = \frac{1}{6} A_\alpha \omega_{AB} \quad C_{ABC} = \frac{1}{6} \xi \omega_{ABC}, \quad G_{\alpha\beta AB} = F_{\alpha\beta AB}, \quad \text{and} \quad G_{\alpha ABC} = \partial_\alpha \xi \omega_{ABC}. \tag{5.2.16}$$

The form $\omega_{AB}$ describes $G_{ABCD}$ in the bulk, via Hodge duality,

$$G_{ABCD} = \frac{\alpha}{6} \varepsilon_{ABCD}{}^{EF} \omega_{EF} \varepsilon(x^{11}). \tag{5.2.17}$$

It should be a member of the cohomology group $H^{2,2}(K_6)$ of the Calabi–Yau. This cohomology class is non-trivial for Calabi–Yau manifolds. The three-form $C_{\alpha\beta\gamma}$ is Hodge dual to a scalar in five dimensions, so it can be replaced by a scalar field $\sigma$, which is

considerably easier to work with.[7] Taking all this into account, the field content we have arrived at consists of a gravity multiplet $(g_{\alpha\beta}, A_\alpha)$ which is an Einstein-frame graviton, a vector, plus fermionic terms which we do not write explicitly. There is also a matter multiplet $(V, \sigma, \xi, \bar{\xi})$, containing the modulus $V$, the scalar fields $\sigma$, $\xi$, $\bar{\xi}$. All of this is to be supplemented with the boundary theories. These theories will be replaced by an arbitrary cosmological fluid when we come to consider phenomenology in Part 2. The total action in the bosonic part consists of a gravitational sector,

$$S_{\text{grav}} = -\frac{1}{2\kappa_5^2}\int_{M_5}\sqrt{-g}\left[R + \frac{3}{2}F_{\alpha\beta}F^{\alpha\beta} + \frac{1}{\sqrt{2}}\varepsilon^{\alpha\beta\gamma\delta\varepsilon}A_\alpha F_{\beta\gamma}F_{\delta\varepsilon}\right], \tag{5.2.19}$$

which contains a Chern–Simons term in analogy with the eleven-dimensional case. The other terms are just a kinetic term for the vector and standard Einstein gravity. The matter multiplet has action

$$S_{\text{matter}} = -\frac{1}{2\kappa_5^2}\int_{M_5}\sqrt{-g}\Bigg[\frac{1}{2}V^{-2}\partial_\alpha V\partial_\beta V + 2V^{-1}\partial_\alpha\xi\partial^\alpha\bar{\xi} + \frac{1}{24}V^2 G_{\alpha\beta\gamma\delta}G^{\alpha\beta\gamma\delta} + \frac{\sqrt{2}}{24}\varepsilon^{\alpha\beta\gamma\delta\varepsilon}G_{\alpha\beta\gamma\delta}\left(\mathrm{i}[\xi\partial_\varepsilon\bar{\xi} - \bar{\xi}\partial_\varepsilon\xi] + 2\alpha A_\varepsilon\right) + \frac{1}{3}V^{-2}\alpha^2\Bigg], \tag{5.2.20}$$

plus a boundary piece which we ignore since, as we have already described, the boundary $E_8$ Yang–Mills theory it describes will be replaced with something less sophisticated later. Higher derivative terms have been deleted. Writing $G_{\alpha\beta\gamma\delta}$ in terms of its dual $\sigma$, the matter theory can be re-expressed in the rather simpler form (for details, refer to Lukas

[7] This is a convenient shorthand for what is really taking place, because a three-form is manifestly dual to a $(5-3)$-form in five dimensions, or a two-form. However, one should remember that $C_{\alpha\beta\gamma}$ is not gauge invariant by itself, and so cannot enter the action in arbitrary combinations, but only in terms of the field strength $G = \mathrm{d}C$ or (in odd dimensions) in the Chern–Simons combination $C \wedge \mathrm{d}C \wedge \cdots \wedge \mathrm{d}C$ with an appropriate normalization (de Azcárraga and Izquierdo, 1995). (Eleven-dimensional supergravity contains such a Chern–Simons term, and we shall see shortly that the same is true for the compactified five-dimensional supergravity.) Therefore one should really be dualising the form $G$, rather than $C$, which gives a $(5-4)$-form, or one-form. This one-form can be expressed in terms of a potential field $\sigma$, which is the scalar field we are concerned with. The potential form is

$$G_{\alpha\beta\gamma\delta} = \frac{1}{\sqrt{2}}V^{-2}\varepsilon_{\alpha\beta\gamma\delta\varepsilon}\left(\partial^\varepsilon\sigma - \mathrm{i}[\xi\partial^\varepsilon\bar{\xi} - \bar{\xi}\partial^\varepsilon\xi] - 2\alpha A^\varepsilon\right). \tag{5.2.18}$$

et al. (1999b))

$$S_{\text{matter}} = -\frac{\text{Vol}(K_6)}{2\kappa^2}\int_{M_5}\sqrt{-g}\left[h_{uv}\nabla_\alpha q^u \nabla^\alpha q^v + \frac{1}{3}V^{-2}\alpha^2\right], \tag{5.2.21}$$

where $q^u$ is the multiplet $(V, \sigma, \xi, \bar{\xi})$ and the covariant derivative $\nabla_\alpha$ acting on the multiplet is defined by the rule

$$\nabla_\alpha q^u = \partial_\alpha q^u + \alpha A_a k^u, \quad \text{where } k^u = (0, -2, 0, 0), \tag{5.2.22}$$

and the sigma-model metric $h_{uv}$ is of Kähler form, $h_{uv} = \partial_u \partial_v K$, and the Kähler potential $K$ satisfies

$$K = -\ln(S + \bar{S} - 2C\bar{C}), \quad \text{where } S = V + \xi\bar{\xi} + \mathrm{i}\sigma \text{ and } C = \xi. \tag{5.2.23}$$

The predictions for the five-dimensional couplings which arise from this theory are

$$\kappa_5^2 = \frac{\kappa^2}{\text{Vol}(K_6)} \quad \text{and} \quad \alpha_{\text{GUT}} = \frac{\kappa^2}{2\,\text{Vol}(K_6)}\left(\frac{4\pi}{\kappa}\right)^{2/3}. \tag{5.2.24}$$

This is the usual situation in Kaluza–Klein compactifications and should be compared with (for example) the prediction from the brane compactification (Eq.(5.1.10) and Eq. (5.1.12)). In Kaluza–Klein compactification, the reduced gravitational coupling is diluted by a factor of the volume of the compact dimension. In the Randall–Sundrum compactification, one can tune the reduction based on the fact that there is an anti-de Sitter cosmological constant in the bulk, and one is dealing with a warped compactification. It is this mechanism which allows Randall and Sundrum to evade the usual arguments that the compact Calabi–Yau dimension must be of order the string or Planck scale, and contemplate the possibility of large extra dimensions (Antoniadis et al., 1998; Arkani-Hamed et al., 1998).

**5.2.2. Cosmological solutions and moving branes.** Having completed the reduction to five dimensions, one now seeks cosmological solutions. We adopt the line element

$$\mathrm{d}s_5^2 = a(y)^2\eta_{ab}\mathrm{d}x^a\mathrm{d}x^b + b(y)^2\mathrm{d}y^2 \tag{5.2.25}$$

and suppose that the moduli $V(y)$ (which expresses the volume or breathing mode of the Calabi–Yau field) varies as one moves across the transverse dimension, but on the four-dimensional Minkowski section. The solution is

$$a = a_0\sqrt{H}, \quad b = b_0 H^2, \quad \text{and} \quad V = b_0 H^3, \tag{5.2.26}$$

where the harmonic form $H$ is

$$H = \frac{\sqrt{2}}{3}\alpha|y| + c_0, \tag{5.2.27}$$

and $a_0$, $b_0$ and $c_0$ are constants of integration. $H$ is harmonic on the transverse dimension up to $\delta$-function singularities, as we have come to expect,

$$\triangle_y H = \frac{2\sqrt{2}}{3}\alpha\left[\delta_{\mathrm{D}}(y) - \delta_{\mathrm{D}}(y - \pi\rho)\right]. \tag{5.2.28}$$

These singularities are supported at the location of the branes, so there are two parallel three-branes, as we expect.[8] This is rather similar to a BDEL metric.

However, the advantage of working with a reduction of the full Hořava–Witten scenario is that one is not restricted to looking for static vacuum about which one works in perturbation theory (as we will do with the BDEL metrics throughout Part 2), but instead can investigate dynamical solutions to the theory and study their behaviour. Of particular interest are a class of solutions which, in addition to the two fixed branes we have already studied, carry moving bulk branes (Copeland, Gray, and Lukas, 2001; Copeland, Gray, Lukas, and Skinner, 2002). In this model, the Hořava–Witten vacuum is supposed to contain other branes, which are dynamic D-brane like objects of the sort discussed in Chapter 3 rather than the $E_8$ super-Yang–Mills branes which are necessary end-of-the-world features in the HW model to cancel the gravitino anomaly. These branes descend through the compactification and appear as auxiliary 4-branes in the low-energy world.

There are several features of this model that are deserving of comment.

- The bulk brane cannot move arbitrarily. Instead, the brane starts in the asymptotic past at rest, and is at rest in the asymptotic future. In the interim, there is a single event where the brane can transit across the bulk. The brane is not allowed to oscillate, to move twice, or to reverse its direction. On the other hand, the brane can collide with one of the orbifold fixed planes which bound the transverse dimension. Upon collision, the brane undergoes a 'small instanton transition' (Gray, 2004; Gray and Lukas, 2003; Gray, Lukas, and Probert, 2004; Ovrut, Pantev, and Park, 2000; Witten, 1999c), where the topological field configurations

[8] This solution has $a^2 \propto |y|$, whereas the BDEL branes have $a^2 \propto \cos + \sin$, which are the appropriate harmonic functions on a circle. There is no inconsistency here; the point is that in the BDEL compactification there is a bulk cosmological constant which is not present in this case. See, eg. Binetruy et al. (2000a) for an explicit comparison. In the case where the bulk is de Sitter rather than anti-de Sitter, so that $\Lambda$ is negative (with our sign conventions), one obtains $a^2 \propto \cosh + \sinh$ instead.

describing the brane (it can considered as a soliton) dissolve and reappear after a phase transition in the boundary super-Yang–Mills theory as a field configuration preserving the conserved charges of the bulk brane.

Generically it is very difficult to follow the phase transition in detail. Instead, the transition is usually phenomenologically parametrized. Very recently, an explicit example of a moduli driven phase transition of this sort has been constructed (Gray, 2004).

- The motion of the brane drives the model to strong coupling, where the radius of the transverse dimension is large (see also Witten (1996)). Therefore the model ends up in the *strongly* coupled state, rather than close to the perturbative $E_8 \times E_8$ heterotic string. This is important, because it shows that the orbifold does not generically contract to zero size, as it might do in a model in which the two orbifold planes themselves approach one another (cf. the discussion of the Ekpyrotic and cyclic models in the next section), but rather the opposite occurs. Therefore the late-time M-theory vacuum would be a strong coupled theory, and not at all close to the perturbative heterotic string. This may go some way to explaining which none of the known perturbative string theories accurately describe the real world.
- It is the interior five-brane, which in terms of the low-energy supergravity is built out of topologically non-trivial field configurations, which moves, and not either of the end-planes which support the $E_8$ super-Yang–Mills theory.

## 5.3. Ekpyrotic and cyclic models

An alternative and rather more ambitious scenario in 2001 by Khoury, Ovrut, Steinhardt and Turok (Khoury, Ovrut, Seiberg, Steinhardt, and Turok, 2002a; Khoury, Ovrut, Steinhardt, and Turok, 2001, 2002b; Steinhardt and Turok, 2001, 2002). In this model, which is also based, in a loose sense, on moving branes, the origin of the hot Big Bang phase of the universe's evolution is identified with a brane collision in the very universe, where the two orbifold fixed planes collide. In this section, we describe only the cyclic model which is both more recent are more fully developed.

The cyclic model is not only a particular model of a brane compactification (though without the kind of thoroughly detailed derivation from low energy M-theory which is available for the heterotic M-theory compactification described in the previous section)

but an entire replacement scenario for cosmology, from very high energy epochs in the distant past to the comparatively cold, low-energy world we see around us today. As such it relies on a very large extrapolation, that the M-theory vacuum corresponding to the Hořava–Witten construction remains a valid picture even to high energies. Recall that the scenario was built only involving low-energy supergravity fields, but in principle should include other components corresponding to the high excitation modes of the string.[9]

In the Hořava–Witten model, the background is entirely supersymmetric and therefore static. It is a BPS state,[10] rather like the Papapetrou–Majumdar metric (Chandrasekhar, 1983) in which the force exerted on one brane by the other owing to exchange of virtual excitations exactly cancels between the bosonic and fermionic sectors. Such states are highly special. In the case of the braneworld, a BPS state implies that the branes are exactly parallel and static. The authors of the cyclic scenario work from the assumption that in Nature, supersymmetry is broken,[11] and propose an ad-hoc potential between the branes. This potential is generically supposed to look something like the Lennard–Jones 6/12 potential which is familiar from elementary studies of molecular bound states,

$$V_{\text{L–J}} = \varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right]. \tag{5.3.1}$$

As the brane separation approaches zero, say within a string length of zero, the potential asymptotes to zero exponentially from below. As the separation diverges, the potential acquires a small positive value. (The behaviour as the separation goes to infinity is not well specified: in the original model, the potential was supposed to approximate a small positive plateau, but it could just as well diverge like a polynomial or faster. But, see Lehners and Stelle (2003).)

---

[9] Of course, this objection is not fatal, and it does not apply only to the cyclic model. The Standard Model relies on the similarly fantastic extrapolation that gravitational physics which has only really been tested at scales of order the size of the solar system (and which could be argued to fail at scales of order the size of galaxies, although other more conservative explanations, such as the introduction of cold dark matter, seem rather more realistic). In the case of M-theory, however, it seems doubtful whether the mathematics we have available to describe the low-energy limit is even capable of writing down what the theory looks like at high energy, so the problem could be expected to be more severe.

[10] In the sense of Bogo'molyni, Prasad and Sommerfeld, who constructed the prototypical BPS monopole.

[11] Supersymmetry, if it exists in nature at all, is certainly broken in our own vacuum.

When the branes are far from each other, they are on the small positive plateau, and there is a small positive vacuum energy density that is supposed to correspond to the effective $\Lambda$ we see in the world today. This plateau is very flat, but not exactly flat, since the configuration is not a BPS state, so the orbifold rolls down the potential very slowly, and the branes approach each other on very long timescales. The orbifold rolls in this potential almost indefinitely, and although energy is expended in each cycle extra energy is supposed to be supplied by the gravitational field. In this respect, the scenario is rather similiar to the old-fashioned Steady State Cosmology of Hoyle, and Bondi & Gold (Hoyle et al., 2000), where in Hoyle's original formulation baryons were steadily created in the space between the galaxies by a notional $c$-field.

The branes collide at regular intervals, and the authors identify the collision as the origin of the high energy hot Big Bang phase. However, there is no real technical innovation here and the comments made above in relation to the heterotic compactification still apply: it is generically very difficult to follow any model through the phase transition that is associated with the a brane collision (but in the cyclic/Ekpyrotic case see Turok, Perry, and Steinhardt (2004)). This is quite bad enough in the case of heterotic compactification, where the orbifold remains at finite size and only a brane formed out of bulk gauge and matter fields is being dissolved. In the cyclic scenario, the entire orbifold shrinks to zero size. At this stage, a couple of comments are appropriate.

(1) This is exactly the scenario which was disfavoured by the heterotic compactification, which found that generically solutions are pushed to strong coupling asymptotically, and remain there.

(2) From the point of view of the low-energy world, something very singular happens when the orbifold shrinks to zero size, because one dimension is momentarily disappearing. However, when lifted ('oxidized') to an M-theory interpretation, the situation appears *a priori* much less confusing. If this scenario does indeed reduce at low energy to M-theory on $M_{10} \times \mathbf{S}^1/\mathbf{Z}_2$, then one should just be approaching the heterotic $E_8 \times E_8$ perturbative string as the transverse dimension is shrunk away. This is a perfectly well behaved string vacuum with no odd effects, or, indeed, bouncing behaviour. So in this case there doesn't seem to be any well-motivated reason to expect that anything special happens after the collision, and since the $E_8 \times E_8$ heterotic string does not appear to be our world, the result is

disappointing. One can evade this negative outcome be supposing that one is in an M-theory vacuum which is not Hořava–Witten M-theory, but then the model loses some of its appeal.

(3) Lehners and Stelle (2003) show that the close approach of the branes is described by a sigma-model, and generically show that there is a hump in the potential. One (at least) has to add an extra assumption, that the orbifold has enough kinetic energy to get over the hump, in order to get the branes to actually collide.

Assuming that the collision actually happens, and that the transition through the singular state has the intended properties, the heat from the aftermath of the collision dominates the energy density of the universe. This corresponds to the radiation dominated and matter dominated epochs of the evolution of our own universe. Inevitably, however, the energy density from any matter and radiation or other non-exotic forms of matter with $\omega > -1/3$ is inexorably redshifted away gravitationally by the expansion of the brane, leaving only an immutable contribution from the inter-brane potential, which is not sitting on the small positive plateau again. Since the brane is effectively carrying a de Sitter cosmology, Wald's no-hair theorem applies, and any large scale structure, anisotropy or vorticity on the brane will decay away exponentially to leave a pair of almost exactly parallel, vacuum, BPS branes. Since the effect is not perfect, however, the orbifold eventually rolls down the plateau again to begin the next cycle of the cosmology.

As the branes roll together, the expansion reverses sign and becomes a collapse. As the branes approach, quantum fluctuations in the location of the brane[12] cause the collision to occur at slightly different places at different points on the brane. Just as in the Guth–Pi formulation of inflationary structure formation, the result is that the matter and radiation

[12]In this sense, the branes are considered to be rather more like the dynamical D-branes of Type II string theories than the orbifold fixed planes of the Hořava–Witten model. Of course orbifold fixed planes are perfectly sensible ideas in string theory, and the theory does have ways of making sense out of these apparently singular quotient-space backgrounds (Greene, 1997). On the other hand, strings which becomes trapped at orbifold fixed points do not appear to move off, so transverse excitations are not possible (Polchinski, 1998). D-branes themselves do not suffer from this difficulty. The transverse oscillations, or collective dynamics, are described by gauge field theories, and there are other types of excitation such as BIons which deform the brane (Johnson, 2003). There doesn't appear to be any reason embedded in the cyclic model (or the older Ekpyrotic model) why the branes have to be orbifold fixed planes, except the desire to connect with Hořava–Witten theory.

that are assumed to created in the aftermath of the brane collision appear at slightly different times, and evolve essentially as separate universes (Liddle and Lyth, 2000; Wands et al., 2000). It is claimed that a scale-invariant spectrum of fluctuations results (Gratton, Khoury, Steinhardt, and Turok, 2004; Khoury et al., 2002b; Khoury, Steinhardt, and Turok, 2003; Tolley, Turok, and Steinhardt, 2004). This claim is somewhat controversial (Allen and Wands, 2004; Brandenberger and Finelli, 2001; Durrer, 2001a; Durrer and Vernizzi, 2002; Lyth, 2002a,b), since some authors claim that the generated spectrum is too steep to accommodate observations. For the present, it seems safest to say that the issue is not yet closed. Notice that in common with all other brane scenarios, the cyclic model involves timelike branes. Therefore the issue of causality is rather subtle, since one is not really beginning with the intial value problem as it is usually formulated in general relativity (Hawking and Ellis, 1973; Seery, 2001; Wald, 1984). Therefore the ekpyrotic collision takes place at all time equally, and the perturbation spectrum is laid down everywhere in the universe, and at all times, at the moment of the collision.

## 5.4. Verlinde compactification

The rest of this chapter is concerned entirely with the BDEL compactifications rather than compactifications of heterotic M-theory or the Ekpyrotic or cyclic scenarios. The ultimate aim is a treatment of bulk quantum fields in BDEL compactifications, but before moving on we pause to give two alternative treatments of the BDEL model. The first of these is based on M-theory and the renormalization group,[13] rather than general relativity or low-energy field theory, and provides a highly conceptual reformulation of the model. This is the method of Verlinde compactification outlined in the present section. In the next section we give what amounts to a different account of the general relativistic formulation, based on a projection of the bulk Einstein equations onto the brane.

The theories outlined above all share a common feature, that the extra fifth dimension which is transverse to the brane really descends from the extra Hořava–Witten dimension and is related to the string coupling. The various models correspond, more or less, to vacua of eleven dimensional supergravity with some additional matter fields, taking the boundary planes where our universe is supposed to live into account. As such, these scenarios bear

[13]This section depends for full understanding on the description of the renormalization group in Section D.1.

a striking similarity to a recent proposal in string theory, known as the AdS/CFT correspondence, which extends the stringy dualities already discussed in Chapter 3. AdS/CFT relates four-dimensional Yang–Mills theory (or, in fact, its $\mathcal{N} = 4$ supersymmetric extension, although we will not require technical details in this section), and Type IIB string theory on the background $\mathrm{AdS}_5 \times \mathbf{S}^5$, where $\mathrm{AdS}_5$ is five-dimensional anti-de Sitter space. This background has the ten-dimensional metric

$$\mathrm{d}s_{10}^2 = \mathrm{e}^{-2y/R}\mathrm{d}s_4^2 + \mathrm{d}y^2 + R^2\mathrm{d}\Omega_5^2. \tag{5.4.1}$$

Here $\mathrm{d}s_4^2$ is a flat four-dimensional Lorentzian metric, and the curvature radius $R$ is

$$R^2 = \alpha'\sqrt{4\pi N g_s} \tag{5.4.2}$$

in which $g_s = g_{\mathrm{YM}}^2$ is the Type IIB string coupling, and $g_{\mathrm{YM}}^2$ is the coupling constant of the Yang–Mills theory (Aharony, Gubser, Maldacena, Ooguri, and Oz, 2000; Johnson, 2003; Maldacena, 1998, 2003b). The $\mathbf{S}^5$ component is unimportant for the discussion which follows. Under the correspondence the coordinate $y$, although a general spacetime dimension, should be thought of as parametrizing the four-dimensional scale. Any two excitations in the SYM theory which are related by a scale transformation $\lambda$, of the form

$$x_4 \mapsto \mathrm{e}^{\lambda}x_4 \tag{5.4.3}$$

translate in AdS metric to two excitations concentrated around different $y$-locations related by a transformation

$$y \mapsto y + \lambda R. \tag{5.4.4}$$

Thus large gauge theory excitations (the far infra-red in field theory terms) correspond to $\lambda \to \infty$, or extremely large $y \to \infty$. Short scale excitations (the field theory ultra-violet) correspond to $y \to -\infty$. The gauge theory itself can be thought of as living on the AdS boundary, and the AdS/CFT correspondence provides a holographic projection of gauge theory physics onto the AdS metric. In detail, the content of the correspondence asserts that given some field $\phi$, propagating on AdS space, with $\phi \to \phi_0$ approaching the boundary, the partition functions of the bulk theory and the boundary CFT are equal,

$$Z_{\mathrm{AdS}}(\phi_0) = Z_{\mathrm{CFT}}(\phi_0). \tag{5.4.5}$$

The bulk theory $Z_{\mathrm{AdS}}$ includes an integration over all fields $\phi$ coinciding with the value $\phi_0$ at the boundary. The boundary theory $Z_{\mathrm{CFT}}$ is the CFT partition function coupled to $\phi_0$ as a source for the CFT operator which corresponds to the bulk field $\phi$ (Johnson, 2003).

Evidently, this form of the AdS metric is equivalent to the Randall–Sundrum model, except that the range of $y$ is finite or semi-infinite. In view of the understanding that $y$ parametrizes the scale of the four-dimensional theory – in short, $y$ is the renormalization group scale for this theory (see Appendix D) – it is clear that truncating the AdS theory to $y$ values less than some fixed point $y_{\mathrm{IR}}$ corresponds to an infra-red cutoff in the gauge theory, whereas demanding that $y$ be greater than $y_{\mathrm{UV}}$ introduces an ultra-violet cutoff. In the full theory, the range of $y$-values extends over the entire real axis, and since the Type II string theory propagating on the AdS space contains closed strings that carry gravitational excitations there are AdS fields that correspond to gravitational excitations. These gravitational excitations are visible in the low-energy theory as supergravity fields on AdS, but the CFT on the boundary does not contain any gravity. This is because modes of the AdS gravitational field that extend to the horizon are not normalizable Breitenlohner and Freedman (1982a,b); Mezincescu and Townsend (1985), and therefore do not fluctuate: they are 'locked'. In the Randall–Sundrum case the range of $y$ is truncated. (In the Kaloper–Linde model, the $y$ range is finite, corresponding to both an infra-red and ultra-violet cutoff; the Randall–Sundrum model extends to $y \to \infty$, and so possesses an ultra-violet cutoff but includes arbitrarily long wavelength excitations.) As an immediate consequence, the AdS theory contains normalizable graviton modes, and the boundary

theory on the brane is deformed to include gravity (Giddings, Katz, and Randall, 2000; Gubser, 2001; Perez-Victoria, 2001; Witten, 1999a).[14]

Since $y$ parametrizes the renormalization group scale, features in the metric as one varies $y$ correspond to non-trivial renormalization group behaviour such as phase transitions or symmetry breaking. In the gravity dual, these features would be visible as domain walls or topological defects. On the other hand, the $y$-dependence of the Randall–Sundrum model is featureless, and decays to zero in the far infra-red. This simply corresponds to the fact that coupling to gravity is energy–dependent, and stronger at high energies (Verlinde, 2000).

**5.4.1. The hierarchy problem.** It is now fairly simple to understand how extra-dimensional scenarios of this sort are associated with attempts to solve the hierarchy problem, which is fundamentally connected with the existence of vastly different energy scales in physics, namely the electroweak scale $M_{\rm EW} \sim 1\,\text{TeV}$ and the Planck scale $M_{\rm P} \sim 10^{19}\,\text{GeV}$.[15] A mass $m$ on the Randall–Sundrum brane at $y = 0$ is related via renormalization group flow to a mass $m_4$ at some other location $y > 0$,

$$m_4(y) = m\mathrm{e}^{-y/\ell}. \tag{5.4.6}$$

As we have explained, increasing $y$ is the infra-red direction of RG flow, so moving to larger $y$ reduces the mass. The solution to the hierarchy problem suggested by Randall &

[14]The appearance of Einstein gravity on the brane as an effective theory of the low-energy physics was something of a surprise, since in classical Kaluza–Klein theories compactified gravity does not become four-dimensional but instead retains signatures of its higher dimensional origin. Precision tests of gravity which are usually interpreted as evidence of only four large dimensions would then rule out extra-dimensional theories, unless the extra dimensions were truly tiny. In part Randall and Sundrum (1999a) was designed to point out this Kaluza–Klein behaviour does not necessarily occur.

Besides the general argument outlined above, the appearance of Einstein gravity has an attractive explanation in terms of the AdS/CFT correspondence, in which the zero mode corresponding to the four-dimensional graviton appears as a result of counterterms which must be added to the theory to render it finite (Hawking, Hertog, and Reall, 2000; Perez-Victoria, 2001). The quadratic correction to the Friedmann equation should be viewed in this sense as a consequence of the trace anomaly of the CFT.

[15]Alternatively one could choose the SUSY scale for the low energy physics, and the string scale or the GUT scale instead of the Planck mass. However the point we wish to make is representative, not detailed, and the GUT, string and Planck scales are all more or less related (Witten, 1996), whereas the putative scale of any low-energy SUSY breaking would determine the masses appearing in the electroweak theory.

Sundrum (Randall and Sundrum, 1999b) simply involves postulating that such hierarchies are generated via (5.4.6), in which Planck scale processes occur near $y = 0$, somewhat distant from out own location, whereas electroweak processes we see in the low energy world are generated near us.[16] Because the RG flow described by varying $y$ is exponential, one does not need large $\Delta y$ to generate large mass hierarchies. By choosing $\Delta y$ appropriately, one can select any attractive scale for the fundamental Planck scale. For example, a popular choice has been to suppose $M_{\rm P} \sim 1\,\text{TeV}$ in order to bring the scale of stringy or quantum gravitational physics down to accelerator energies. Having done so, one is reliant on precision tests of gravity (such as the Eötvös or Cavendish experiments (Weinberg, 1972) to rule out the possibility of such large extra dimensions. At present it seems that 1 TeV is a little low for the fundamental scale, although the evidence is not yet conclusive.

## 5.5. The brane Einstein equations

The discussion of the Einstein field equations for the metric BDEL compactification described at the top of this chapter centred on field equations for the metric fields $a$ and $n$. Once these fields are known, we are entitled to restrict attention to the brane slice at $y = 0$ in order to find the geometry on the brane. Of course, there is nothing wrong with this approach, but it is important to realise that the method of working is fundamentally different from what we have come to expect on the basis of four-dimensional general relativity. In four dimensions we have the familiar Einstein field equation $G_{ab} = \kappa^2 T_{ab}$, which allows us to prescribe any physically sensible matter distribution $T_{ab}$ (Hawking and Ellis, 1973) and then calculate the behaviour of the gravitational field. In particular, this allows us to address global questions such as the positivity of energy (Perry, 1984). This depends on *global* properties (Flaherty, 1984), such as knowing the rate of fall-off of the gravitational field near infinity produced by a given mass distribution. Global information of this sort is most easily obtained by a careful study of the Einstein equations. In addition, comparison with four-dimensional gravity is facilitated by comparison of the

[16] One can use a somewhat similar effect to try and obviate the necessity for inflation to generate initial perturbations at some prescribed (small) scale by 'lensing' perturbations at a distant point onto the brane (Chung and Freese, 2003). The attraction here is that the fine-tuning conditions on inflation may be relaxed if the parameter window it must hit to generate phenomenological (or 'anthropically') attractive models can be expanded. However since inflation must always happen on our brane, it is not clear whether one gains anything from this kind of argument.

four-dimensional effective field equations. For all of these reasons, not to mention the elegance of the procedure, it is desirable to attempt to obtain a set of effective four-dimensional Einstein equations for the braneworld. This programme was first accomplished by Shiromizu, Maeda, and Sasaki (2000) and is reviewed in this section.

Notice that this is only a reformulation of five-dimensional general relativity, and as such contains no new information. Instead, the four-dimensional projected Einstein equations simply represent the same physics in a more convenient form for use when proving global theorems.

The starting point is Gauss' equation (1.3.11) for the Ricci tensor induced on the brane by the bulk geometry. For a timelike brane with spacelike unit normal $n_a$, this reads

$$R_{ab} = h^e{}_a h^g{}_b \mathsf{R}_{eg} - n^f n_k h^e{}_a h^g{}_b \mathsf{R}^k{}_{efg} + \chi_{ab}\chi - \chi_{ae}\chi^e{}_b, \tag{5.5.1}$$

where the brane tensor $h_{ab} = g_{ab} - n_a n_b$ is the induced metric, or first fundamental form, and $\chi_{ab} = h_a{}^c h_b{}^d \nabla_c n_d$ is the extrinsic curvature, or second fundamental form. For the purposes of this section, we are adopting a notation in which four-dimensional quantities are denoted in conventional italic type, but five-dimensional quantities, such as the five-dimensional Riemann curvature $\mathsf{R}_{abcd}$ are denoted in a sans-serif face. In contracted form, this is an expression for the Ricci curvature,

$$R = \mathsf{R} - 2\mathsf{R}_{ab} n^a n^b + \chi^2 - \chi^{ab}\chi_{ab}. \tag{5.5.2}$$

The Einstein tensor on the brane is

$$\begin{aligned} G_{ab} &= R_{ab} - \frac{1}{2} R h_{ab} \\ &= h^e{}_a h^g{}_b \left( \mathsf{R}_{eg} - \frac{1}{2}\mathsf{R} g_{eg} \right) + h_{ab} n^e n^g \mathsf{R}_{eg} - \mathsf{E}_{ab} + \chi_{ab}\chi - \chi_{ae}\chi^e{}_b - \frac{1}{2} h_{ab} \left( \chi^2 - \chi^{eg}\chi_{eg} \right). \end{aligned} \tag{5.5.3}$$

where the tensor $\mathsf{E}_{ab}$ is given by

$$\mathsf{E}_{ab} = n^f n^k h^e{}_a h^g{}_b \mathsf{R}_{kefg}. \tag{5.5.4}$$

The point of separating this combination from the remaining terms is that it can be rewritten in terms of the conformal or Weyl tensor, $C_{abcd}$, which is defined by the rule Hawking and Ellis (1973)

$$C_{abcd} = R_{abcd} + \frac{1}{n-2}\left( g_{ad}R_{bc} - g_{ac}R_{bd} + g_{bc}R_{ad} - g_{bd}R_{ac} \right) + \frac{1}{(n-1)(n-2)} R (g_{ac}g_{bd} - g_{ad}g_{bc}). \tag{5.5.5}$$

If we define a tensor $\mathsf{F}_{ab}$ as a combination of the Weyl tensor and the same contractions as $\mathsf{E}_{ab}$, so that $\mathsf{F}_{ab} = n^f n^k h^e{}_a h^g{}_b \mathsf{C}_{kefg}$, then we have the simple relation

$$\begin{aligned}\mathsf{F}_{ab} = \mathsf{E}_{ab} + \frac{1}{3}\Big( & h^e{}_a h^g{}_b n^f n_k \delta^k_g \mathsf{R}_{ef} - h^e{}_a h^g{}_b n^f n_k \delta^k_f \mathsf{R}_{eg} + h^e{}_a h^g{}_b n^f n^k g_{ef} \mathsf{R}_{kg} \\ & - h^e{}_a h^g{}_b n^f n^k g_{eg} \mathsf{R}_{kf}\Big) + \frac{1}{2}\mathsf{R}\left(h^e{}_a h^g{}_b n^f n_k \delta^k_f g_{eg} - h^e{}_a h^g{}_b n^f n_k \delta^k_g g_{ef}\right).\end{aligned} \tag{5.5.6}$$

Half of the terms appearing here vanish owing to the orthogonality of $h_{ab}$ and $n^a$. These are the terms where an $n$ is contracted with $h$. This reduces the remaining pieces to

$$\mathsf{F}_{ab} = \mathsf{E}_{ab} - \frac{1}{3}\left(h^e{}_a h^g{}_b \mathsf{R}_{eg} + h_{ab} n^e n^g \mathsf{R}_{eg}\right) + \frac{1}{12}\mathsf{R} h_{ab}. \tag{5.5.7}$$

Now replacing $\mathsf{E}_{ab}$ by an equivalent expression in terms of $\mathsf{F}_{ab}$ in (5.5.3) gives

$$\begin{aligned}G_{ab} = {} & h^e{}_a h^g{}_b \left(\mathsf{R}_{eg} - \frac{1}{2}\mathsf{R} g_{eg}\right) - \mathsf{F}_{ab} - \frac{1}{3}\left(h^e{}_a h^g{}_b \mathsf{R}_{eg} + h_{ab} n^e n^g \mathsf{R}_{eg}\right) \\ & + \frac{1}{2} h_{ab}\mathsf{R} + h_{ab} n^e n^g \mathsf{R}_{eg} + \chi_{ab}\chi - \chi_{ae}\chi^e{}_b - \frac{1}{2} h_{ab}\left(\chi^2 - \chi^{eg}\chi_{eg}\right).\end{aligned} \tag{5.5.8}$$

The dependence of the four-dimensional Einstein tensor on the matter content of the theory can be made explicit through a re-writing of the five-dimensional curvature quantities in terms of the energy–momentum tensor of the bulk. The Einstein equations in the bulk are

$$\mathsf{R}_{eg} - \frac{1}{2}\mathsf{R} g_{eg} = \kappa_5^2 \mathsf{T}_{eg}. \tag{5.5.9}$$

Contracting on $e$ and $g$ gives an expression for the five-dimensional Ricci curvature,

$$\mathsf{R} = -\frac{2\kappa_5^2}{3}\mathsf{T}. \tag{5.5.10}$$

The five-dimensional Ricci tensor can be written as just a function of the energy–momentum tensor alone, without involving $\mathsf{R}$,

$$\mathsf{R}_{eg} = \kappa_5^2\left(\mathsf{T}_{eg} - \frac{1}{3} g_{eg}\mathsf{T}\right). \tag{5.5.11}$$

Combining these expressions allows us the rewrite the four-dimensional Einstein tensor in terms of the five-dimensional matter content,

$$\begin{aligned}G_{ab} = {} & \kappa_5^2 h^e{}_a h^g{}_b \mathsf{T}_{eg} - \frac{\kappa_5^2}{3} h^e{}_a h^g{}_b \left(\mathsf{T}_{eg} - \frac{1}{3} g_{eg}\mathsf{T}\right) - \frac{\kappa_5^2}{3} h_{ab} n^e n^g \left(\mathsf{T}_{eg} - \frac{1}{3} g_{eg}\mathsf{T}\right) \\ & - \frac{\kappa_5^2}{18} h_{ab}\mathsf{T} + \kappa_5^2 h_{ab} n^e n^g \left(\mathsf{T}_{eg} - \frac{1}{3} g_{eg}\mathsf{T}\right) + \chi_{ab}\chi - \chi_{ae}\chi^e{}_b - \frac{1}{2} h_{ab}\left(\chi^2 - \chi^{eg}\chi_{eg}\right) - \mathsf{F}_{ab}.\end{aligned} \tag{5.5.12}$$

At this point, collecting terms allows us to write this more simply as

$$G_{ab} = \frac{2\kappa_5^2}{3}\left[h^e{}_a h^g{}_b \mathsf{T}_{eg} + h_{ab}\left(n^e n^g \mathsf{T}_{eg} - \frac{1}{4}\mathsf{T}\right)\right] + \chi_{ab}\chi - \chi_{ae}\chi^e{}_b - \frac{1}{2}h_{ab}\left(\chi^2 - \chi^{eg}\chi_{eg}\right) - \mathsf{F}_{ab}. \tag{5.5.13}$$

There is also Codacci's equation (1.3.8), which constrains the derivatives of the extrinsic curvature,

$$\nabla_a \chi^a{}_b - \nabla_b \chi = \mathsf{R}_{eg} h^e{}_b n^g = -\kappa_5^2 h^e{}_b n^g \mathsf{T}_{eg}. \tag{5.5.14}$$

So far this has all been completely general and would apply for any embedded surface. In order to relate this formalism to the metric branes discussed above we fix Gaussian normal coordinates (5.1.1), in terms of which $n = \mathrm{d}y$ and the metric in the bulk

$$\mathrm{d}s^2 = h_{ab}\mathrm{d}x^a \mathrm{d}x^b + \mathrm{d}y^2. \tag{5.5.15}$$

The extrinsic curvature of the brane is

$$\chi_{ab} = (\delta_a^c + n^c n_a)(\delta_b^d + n^d n_b)\nabla_c n_d = \Gamma^y_{ab} = -\frac{1}{2}\partial_y h_{ab}. \tag{5.5.16}$$

This is a general result for the extrinsic curvature in Gaussian normal coordinates.[17] If the energy–momentum tensor in the bulk consists of the vacuum contribution plus the energy–momentum carried by the brane, so that $\mathsf{T}_{ab}$ satisfies

$$\mathsf{T}_{ab} = \Lambda g_{ab} + S_{ab}\delta_{\mathrm{D}}(y), \tag{5.5.18}$$

and we separte the brane contribution into a tension piece plus matter, $S_{ab} = -\lambda h_{ab} + \tau_{ab}$ (but see Bowcock et al. (2000)), then $\tau_{ab}$ is a tensor on the brane which can be identified with the matter theory it is carrying.[18] The Israel junction condition (see Chapter 1; Israel (1966)) produces a discontinuity in the geometry, which in $n$ dimensions says

$$[h_{ab}]^+_- = 0 \quad \text{and} \quad [\chi_{ab}]^+_- = -\frac{8\pi}{M_{\mathrm{P}}^{n-2}}\left(S_{ab} - \frac{1}{n-2}h_{ab}h^{cd}S_{cd}\right). \tag{5.5.19}$$

[17] In more detail,

$$\chi_{ab} = \nabla_a n_b + n_b n^d \nabla_a n_d + n^c n_a \nabla_c n_b + n^c n_a n^d n_b \nabla_c n_d, \tag{5.5.17}$$

and since $n_a n^a = 1$ the terms involving $n^d \nabla_a n_d$ vanish. The nerm $n^c \nabla_c n_b = \nabla_y n_b$ is the derivative of $n_b$ along the off-brane direction, which in these coordinates is just locally zero. One can see this formally by writing $\nabla_y n_b = \partial_y n_b + \Gamma^y_{yb} = \Gamma^y_{yb}$, and this connexion component vanishes.

[18] The equation of state for $\lambda$ here is $p = -\rho$, with $\rho$ positive for positive $\lambda$. Therefore $\lambda$ can be interpreted as the tension of the brane; if $\lambda < 0$ it is under tension, whereas is $\lambda > 0$ it is under compression.

Imposing this junction condition for the choice of braneworld coordinates and the brane energy–momentum tensor, taking into account the $\mathbf{Z}_2$ symmetry of the orbifold, shows that

$$\chi_{ab} = -\frac{\kappa_5^2}{2}\left(S_{ab} - \frac{1}{3}h_{ab}h^{cd}S_{cd}\right) = -\frac{\kappa_5^2}{2}\left(\tau_{ab} - \frac{1}{3}\lambda h_{ab} - \frac{1}{3}h_{an}\tau\right). \tag{5.5.20}$$

Notice than when applied to this choice of $\chi_{ab}$, the Codacci equation requires

$$\nabla_a \chi^a{}_b - \nabla_b \chi = 0, \tag{5.5.21}$$

since the right-hand vanishes on this matter content. Working out the derivatives explicitly, it can be shown that this constraint reduces to the requirement of covariant four-conservation of the brane matter theory,

$$\nabla_a \tau^{ab} = 0. \tag{5.5.22}$$

We are now in a position to write down the Einstein equations for the brane. These are found by evaluating the four-dimensional Einstein tensor over a thin $y$-slice centred around the brane at $y = 0$, and taking the limit as this slice decreases to zero width. This subtle limiting procedure is necessary because of the distributional nature of the bulk energy–momentum tensor.

The totally contracted extrinsic curvature and its square satisfy

$$\chi = \frac{\kappa_5^2}{2}\left(\frac{4}{3}\lambda + \frac{1}{3}\tau\right) \quad \text{and} \quad \chi^2 = \frac{\kappa_5^4}{4}\left(\frac{16}{9}\lambda^2 + \frac{1}{9}\tau^2 + \frac{8}{9}\lambda\tau\right). \tag{5.5.23}$$

The other necessary contractions are

$$\chi_{ae}\chi^e{}_b = 2\kappa_5^2\left(\frac{1}{9}\lambda^2 h_{ab} - \frac{2}{3}\lambda\tau_{ab} + \frac{2}{9}\lambda\tau h_{ab} + \tau_{ae}\tau^e{}_b - \frac{2}{3}\tau\tau_{ab} + \frac{1}{9}\tau^2 h_{ab}\right) \tag{5.5.24a}$$

$$\chi^{eg}\chi_{eg} = 2\kappa_5^2\left(\frac{4}{9}\lambda^2 + \frac{2}{9}\lambda\tau + \tau^{eg}\tau_{eg} - \frac{2}{9}\tau^2\right) \tag{5.5.24b}$$

$$\chi_{ab}\chi = 2\kappa_5^2\left(\frac{4}{9}\lambda^2 h_{ab} + \frac{5}{9}\lambda\tau h_{ab} - \frac{4}{3}\lambda\tau_{ab} - \frac{1}{3}\tau\tau_{ab} + \frac{1}{9}\tau^2 h_{ab}\right). \tag{5.5.24c}$$

The five-dimensional quantities are

$$h^e{}_a h^g{}_b \mathsf{T}_{eg} = \Lambda h_{ab}, \quad n^e n^g \mathsf{T}_{eg} = \Lambda, \quad \text{and} \quad \mathsf{T} = 5\Lambda. \tag{5.5.25}$$

Totalling the various contributions, this gives

$$G_{ab} = \Lambda_4 h_{ab} + \kappa_4^2 \tau_{ab} + \kappa_5^4 \pi_{ab} - \mathsf{F}_{ab}, \tag{5.5.26}$$

where the four-dimensional cosmological constant is

$$\Lambda_4 = \frac{\kappa_5^2}{2}\left(\Lambda + \kappa_5^2\frac{\lambda^2}{6}\right), \tag{5.5.27}$$

and the four-dimensional gravitational coupling is

$$\kappa_4^2 = -\kappa_5^4\frac{\lambda}{6}. \tag{5.5.28}$$

(One can absorb some factors of the gravitational couplings into $\lambda$ and $\Lambda$, which is what was done in the metric-field analysis above.)

Notice that the existence of Newton's constant $G = \kappa_4^2/8\pi$ depends on the presence of a tension $\lambda$ on the brane, and that if $\lambda > 0$ then one has the wrong sign. Therefore gravity on the brane can be either attractive or repulsive, depending on whether the tension $\lambda$ is positive or negative. In order to obtain the conventional gravity which which we are familiar, we need the brane to be under tension.

**5.5.1. Cosmological solutions.** If desired, one can apply this formalism to derive the cosmological BDEL solutions discussed above. Since the analysis has already been carried out above we describe this only in very abbreviated form. The brane matter theory is taken to be of the form

$$\tau_{ab} = (p+\rho)u_a u_b + p h_{ab}, \tag{5.5.29}$$

which is the energy–momentum tensor for a perfect fluid of pressure $p$, density $\rho$ and velocity four-vector $u_a$. We assume that the four-dimensional metric is of the general FRW form (where $k \in \{-1, 0, +1\}$ characterizes the curvature of spatial slices)

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a_{\mathrm{b}}^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right), \tag{5.5.30}$$

and $\mathrm{d}\Omega^2$ is the metric on a unit two-sphere. The Friedmann equation arises from the time–time component of the projected Einstein equations, and reads

$$H^2 + \frac{k}{a_{\mathrm{b}}^2} = \frac{\Lambda_4}{3}h_{ab} + \kappa_4^2\rho + \kappa_5^4\frac{\rho^2}{36}. \tag{5.5.31}$$

By completing the square and using the definitions of the coupling constants, this can be massaged into the form which arose from the metric-field theory.[19]

[19] In some applications it is convenient to have an explicit expression for the Raychaudhuri equation to hand, which can be found by taking the time derivative of the Friedmann equation,

$$2H\dot{H} + 2H\frac{k}{a_{\mathrm{b}}^2} = \frac{\kappa_4^2}{3}\dot{\rho}\left(1 + \frac{\rho}{\lambda}\right). \tag{5.5.32}$$

## 5.6. Stability of the brane

Having described the various major families of brane universe, the next step in formalizing a secure scaffold on which to build a cosmology of our universe is to establish their stability. We briefly alluded to problems with the stability of brane worlds when discussing the lack of anti-de Sitter no-hair theorem above. This means it is not possible to prove rigorous results which demonstrate, for example, the end-point in the evolution of all brane cosmologies which carry a non-zero cosmological constant is the exact de Sitter world. This may turn out to be of some considerable relevance to our universe, which appears on the basis of current observations to fall into this category. One might imagine that it is perfectly sensible to try and generalize Wald's four-dimensional argument to a cosmology carried by the brane, but the problem arises when one tries to constrain possible information falling onto the brane from far in the bulk. There is no way to show that bulk perturbations smooth out over time, so any de Sitter state in the asymptotic future could always be disrupted by new material falling onto the brane.

The no-hair theorem forms one half of a pair of important results which buttress our understanding of the vacua of general relativity in four dimensions. The other is the positive energy theorem of Schoen and Yau (1979). This extremely non-trivial result can be understood as a higher-dimensional generalization of the classical singularity theorems of Penrose, Hawking and Geroch (Hawking and Ellis, 1973). In proving these theorems, one argues by contradiction, constructing putative worldlines of hypothetical observers whose existence is contradictory, whereas Schoen and Yau constructed surfaces with similar properties. The proof can be considerably simplified using twistor techniques (Witten, 1981), and in this form can be fairly straightforwardly generalized to anti-de Sitter space (Flaherty, 1984; Perry, 2001) via a supercovariantization argument. There is no analogous extension of the result to de Sitter space. This happens for a good reason: de Sitter space is essentially unstable to the nucleation of black holes, because the path integral has a saddle at the Nariai metric, which is the instanton for black hole production immersed in a de Sitter background (Perry, 1982).

---

Taken together with the conservation equation $\dot{\rho} + 3H(p + \rho) = 0$, this is equivalent to

$$\dot{H} + H^2 = -\frac{\kappa_4^2}{6}\left[(\rho + 3p)\left(1 + \frac{\rho}{\lambda}\right) + \frac{\rho^2}{\lambda}\right]. \tag{5.5.33}$$

The left-hand side is occasionally written as $\ddot{a}/a$.

These results solidify our understanding of the maximally symmetric vacua of general relativity in immediately useful ways. The Schoen–Yau/Witten positive energy results guarantee that Minkowski space cannot decay into some other vacuum state, and the same applies to anti-de Sitter space. De Sitter space may nucleate black holes,[20] but is an attractor state for any universe with positive vacuum energy. Exactly the same concerns occur in the braneworld: one wishes to show that a universe carried on a Minkowski brane, for example, will not decay into some other state, or at least that such decays happen on a timescale much longer than the age of the universe. However, the stringy character of the compactification radically alters our expectation about how much it will be possible to rigorously prove. The current position can be conveniently summarised:

- There is no positive energy theorem even for Minkowski branes. One can solve the asymptotic twistor equation (Stewart, 1991; Witten, 1981) to reproduce Witten's spinorial argument from four dimensions, but the result for the mass at infinity is not positive definite.[21] This happens because the strong energy condition, which guarantees positivity in four dimensions, is not sufficiently strong to govern the behaviour of the quadratic tensor $\tau_{ab}$ which appears in the brane Einstein equations, and, moreover, the tensor $\mathsf{E}_{ab}$ is not constrained at all. (However for BPS brane configurations in a supergravity vacuum, such as the theory of supersymmetric Randall–Sundrum branes, one expects positivity results in the normal sense; see Bergshoeff, Kallosh, and Van Proeyen (2000, 2001).)
- The no-hair theorem can be generalized to the brane (Santos, Vernizzi, and Ferreira, 2001), although in slightly weakened form. In particular, any strong

[20]This may be of some importance during a vacuum dominated epoch of the Universe's evolution. However, the absence of rigorous stability results should be interpreted as a reflection of the peculiarity of de Sitter space within string theory (Kachru et al., 2003b) as a resonant state which must eventually decay into some other spacetime in which future null infinity is not spacelike.

[21]This calculation does not seem to have appeared in the literature. However it is a trivial modification of the argument from four dimensions. No use is made of the Einstein equations in Witten's argument until it is necessary to replace the curvature tensor $R_{ab}$ in terms of the energy and momentum carried by the space time. This is the crucial step which fails, because of the tensor $\mathsf{F}_{ab}$, which cannot be predicted on the basis of data given on the brane.

anisotropic stress may cause the brane to collapse. However, for sufficiently sanitised inflationary models the homogeneous, isotropic de Sitter world is still the end-point of inflationary evolution.

- The situation regarding black holes on the braneworld is presently undecided (Chamblin, Hawking, and Reall, 2000; Guedens, Clancy, and Liddle, 2002; Gutowski and Reall, 2004; Kudoh and Wiseman, 2004; Reall, 2003). In particular, an exact black hole metric on the brane is not yet known. However, there is no reason to believe that de Sitter space on the brane will not be unstable to black hole nucleation in the same way as in four dimensions. This is an issue of some considerable interest, and we will return to it briefly at the end of Chapter 9.

**5.6.1. Tachyon matter and brane instabilities.** The conclusion from the absence of rigorous stability results is that brane vacua may generically exhibit instabilities of one form or another. The reason for this is not hard to find, and arises from the observation that branes are dynamical objects in string theory. The exchange of string modes between branes may cause attractive or repulsive forces to change the configuration. This need not be a bad thing: the cyclic and Ekpyrotic scenarios, for example, rely on this very mechanism. In simple cases exchange of virtual string modes may only cause the various moduli fields which parametrize the vacuum, such as the inter-brane distance, to roll. In more complicated scenarios the vacuum itself may change as a phase transition occurs which causes a shift to some other vacuum. For example, new branes may condense out of the vacuum or topologically non-trivial field configurations may be generated by the collision of branes (Gray, 2004). Instabilities of this sort are rather generally signalled by the presence of tachyonic fields in the bulk. In particular, an interpretation of the open string tachyon of bosonic string theory strongly advocated by Sen (Sen, 1998, 1999, 2002a,b, 2003) characterizes the vacuum of bosonic string theory as an unstable background built out of branes. The bosonic string tachyon is a field mode which describes the propensity of the branes in this background to annihilate. (See also Frolov, Kofman, and Starobinsky (2002); Gibbons (2003).)

Exactly the same effect can be expected to occur in cosmological brane compactifications Frolov and Kofman (2004). In this case when one includes a bulk scalar field, this field typically displays tachyonic instabilities. There are various ways in which one can attempt to stabilize the vacuum (Goldberger and Wise, 2002), but appeal to the higher

dimensional string theory shows that in the absence of supersymmetry the vacuum is likely to be unstable. We do not address stability issues in this work, ascribing the problem to the general issue of constructing sane string compactifications.

## 5.7. BDEL brane compactifications and zero modes of the graviton

Since the basic brane world geometries, together with their attendant interpretations in terms of stringy physics (or CFTs), and some simple properties like stability have now been established, or at least clarified, the final step is to try and describe low energy physics from the point of view of an observer sitting on one of the branes. The principal low energy excitations are the Kaluza–Klein modes of the supergravity fields in the bulk. In a traditional string theory compactification these Kaluza–Klein modes would be described by the technology outlined in Section 3.2. The metric for the theory, containing gravitational excitations of all types, is

$$\mathrm{d}s^2 = -n^2(1+2\phi)\mathrm{d}t^2 + a^2(\delta_{ij}+e_{ij})\mathrm{d}x^i\mathrm{d}x^j + 2A_i\mathrm{d}x^i\mathrm{d}t + \mathrm{d}y^2 \tag{5.7.1}$$

which consists of a scalar $\phi$, a vector $A_i$, and a tensor $e_{ij}$. We have used some of the gauge freedom to retain Gaussian normal coordinates, meaning that there is no perturbation involving $y$-components of the metric. This is no more than the simple the Kaluza–Klein decomposition which we have already seen, Eq. (3.2.1). The Kaluza–Klein vector $A_i$ and scalar $\phi$ have no zero-modes in the sense of (3.2.5b), but carry a spectrum of massive excitations with mass $m > 3H/2$ (Bridgman, Malik, and Wands, 2002; Frolov and Kofman, 2002).[22] These excitations are not important for inflation, so we ignore $A_i$ and $\phi$ in what follows, concentrating only on the symmetric tensor piece $e_{ij}$.

We will frequently need the curvature quantities corresponding to this metric. For the purposes of this chapter, we are only interested in building the $e_{ij}$ field equation, so

[22] In talking about Kaluza–Klein modes we are being slightly sloppy with notation, since the Kaluza–Klein fields *per se* exist only in the case where the brane world can be given a bundle structure, as described above, or more physically where the wave equation is separable: in physical terms, this means the cosmology carried by the brane is maximally symmetric. This special case is assumed universally throughout the present section. Notice that it is for this reason that we expended so much effort in describing the topological stability of the zero mode when discussing brane world compactifications, because eventually it will be important to consider universes which are marginally perturbed away from maximal symmetry. In this case, the Kaluza–Klein decomposition will not exist and it is no longer clear that a normalizable zero mode persists.

a calculation to first order in $e_{ij}$ will suffice. In later chapters we will be working in the path integral formalism, where the appropriate quantity is not the field equation but the action: in this case, a more careful approach is needed, and one must evaluate the action to second order in the fields. This will be done explicitly in Chapter 6, but would be needlessly troublesome to obtain here. We take the matter content to be the same as (5.1.2), including a bulk source $Q_{ab}$ and a brane energy–momentum tensor $S_{ab}$. The equations governing $a$ and $n$ match the unperturbed case, whereas one can find an evolution equation for $e_{ij}$ by writing down the $(i,j)$ Einstein equation,

$$\frac{a^2}{2}\Box e_{ij} + (\delta_{ij} + e_{ij})Z_0 = \Lambda g_{ij} - \lambda\delta_{\rm D}(y)h_{ij} + \kappa_5^2\left[Q_{ij} - \delta_{\rm D}(y)S_{ij}\right], \tag{5.7.2}$$

where $\Box$ is the *scalar* background Laplacian,

$$\Box = \frac{1}{\sqrt{-g_0}}\partial_a\left(\sqrt{-g_0}g_0^{ab}\partial_b\right), \tag{5.7.3}$$

and $Z_0$ is a background geometrical quantity,

$$Z_0 = a'^2 - \frac{\dot{a}^2}{n^2} + 2aa'' - 2\frac{a\ddot{a}}{n^2} + 2aa'\frac{n'}{n} - 2\frac{a\dot{a}}{n^2}\frac{\dot{n}}{n} + a^2\frac{n''}{n}. \tag{5.7.4}$$

The value of $Z_0$ is set by the background Einstein equations, and is most easily found by picking out the $O(1)$ part of (5.7.2). We find

$$\frac{Z_0}{a^2} = \Lambda - \lambda\delta_{\rm D}(y), \tag{5.7.5}$$

which allows $Z_0$ to be eliminated from the $O(e)$ equation, to give

$$\frac{a^2}{2}\Box e_{ij} = \kappa_5^2(Q_{ij} + \delta_{\rm D}(y)S_{ij}). \tag{5.7.6}$$

Integrating over a small neighbourhood of $y = 0$ shows that $e_{ij}$ obeys the Israel jump condition (Langlois, Maartens, and Wands, 2000)

$$[e'_{ij}]^+_- = -\frac{2n^2}{a^2}S_{ij}. \tag{5.7.7}$$

If there is no anisotropic stress, then $[e'_{ij}]^+_- = 0$. This is a strong restriction on the kinds of boundary conditions which are allowed at the brane, for which tensor fields such as $e_{ij}$ are much more constrained than lower spin fields (Flachi, Moss, and Toms, 2001; Flachi and Toms, 2001).

If there are no sources $Q_{ab}$, $S_{ab}$ then the field equation is $\Box e_{ij} = 0$ and one can construct by hand an effective action,

$$S_g = \frac{1}{8\kappa_5^2} \int \mathrm{d}t\, \mathrm{d}^3x\, \mathrm{d}y\; na^2 \left( \frac{1}{n^2} \dot{e}_{ij} \dot{e}^{ij} - \frac{1}{a^2} \partial^k e_{ij} \partial_k e^{ij} - e'_{ij} e'^{ij} \right). \tag{5.7.8}$$

Up to possible boundary terms or cocycles, this Lagrangian describes the on-shell graviton.[23]

**5.7.1. Canonical quantization.** Let us quantize the field $e_{ij}$ using the canonical procedure. This is fairly awkward, and in any case will not be the preferred method when dealing with the more complicated situations encountered in Part 2. We give the canonical treatment here in order to provide a sound point of departure for the path integral formalism, and so that we can contrast the approaches.

The field canonically conjugate to $e_{ij}$ is

$$\pi^{ij} = \frac{\delta S_g}{\delta \dot{e}_{ij}} = \frac{1}{4\kappa_5^2} \frac{a^3}{n} \dot{e}^{ij}. \tag{5.7.9}$$

The field equation for $e_{ij}$ is the Klein–Gordon equation for the tensor $e_{ij}$, considered as a scalar on spacetime indices. Therefore the effective action must be the Klein–Gordon equation in unusual coordinates, and quantization could be effected by repeating the familiar process of Klein–Gordon quantization on curved backgrounds (Birrell and Davies, 1982; Wald, 1994). This would involve building the conserved Klein–Gordon current $j_a$, which arises from $U(1)$ phase transformations of the field $e_{ij}$, and converting $j_a$ to a conserved inner product. The quantization consists of building the quantum field $\hat{e}_{ij}$ out of the Hilbert space of postive frequency eigensolutions to the field equation equipped with the Klein–Gordon inner product. Fortunately, there is a constructive alternative which essentially consists of the argument of Section 2.1 in reverse (Carlip, 1998; Wald, 1994).

Let $e_\alpha$ be a family of solutions to the field equation $\Box e_\alpha = 0$, with the internal $\mathfrak{so}(3)$ indices omitted for clarity. One can form a symplectic product on these solutions by the rule

$$\omega(e_\alpha, e_\beta) = \int_{\Sigma_t} \mathrm{d}^3x\, \mathrm{d}y\; (e_\alpha \pi_\beta - e_\beta \pi_\alpha) = \frac{1}{4\kappa_5^2} \int_{\Sigma_t} \mathrm{d}^3x\, \mathrm{d}y\; \frac{a^3}{n} (e_\alpha \dot{e}_\beta - e_\beta \dot{e}_\alpha), \tag{5.7.10}$$

where the integration is taken over a spatial hypersurface $\Sigma_t$ (ordinarily a Cauchy surface, but in the anti-de Sitter case there are some subtleties: the ordinary theory of quantization

[23] It is useless when working off-shell, as one does when properly fixing the gauge, to be described in Chapter 6. The action has been artificially gauge-fixed by making $e_{ij}$ transverse and traceless.

on curved manifolds is only guaranteed to work in globally hyperbolic spacetimes (Wald, 1994). Anti-de Sitter space is not globally hyperbolic (Hawking and Ellis, 1973), so there is a difficulty. This problem was resolved and a satisfactory theory of quantization in anti-de Sitter space was worked out in the context of gauged extended supergravity (see Mezincescu and Townsend (1985), and Ceresole and Dall'Agata (2000) four the five-dimensional case relevant to the case at hand) in a series of papers completed in the mid-1980s (Avis, Isham, and Storey, 1978; Breitenlohner and Freedman, 1982a,b; Gibbons, Hull, and Warner, 1983; Mezincescu and Townsend, 1985).[24] The quantization prescription which was developed consists in compactifying AdS and imposing boundary conditions on the solutions to the classical equations of motion entering the Hilbert space of the quantum theory: namely, that the Hilbert space inner product to be constructed below actually is conserved, and that neither particles nor energy are exchanged with the boundary (Balasubramanian, Kraus, and Lawrence, 1999; Klebanov and Witten, 1999; Minces and Rivelles, 2001; Witten, 1998). This prescription was used to understand the Hilbert space of bulk fields in the AdS/CFT correspondence described in Section 5.4 above, and forms the basis for the division of bulk fields into normalizable and non-normalizable modes, the first of which fluctuate and enter the bulk Hilbert space, and the latter of which do not. This is important in AdS/CFT for understanding the appearance of four-dimensional gravity on the brane, as we have described.

Let $\mathcal{S}$ be the space of classical solutions to the equations of motion, and now suppose that the $e_\alpha$ span $\mathcal{S}$. An inner product $(\ ,\ ) : \mathcal{S} \times \mathcal{S} \to \mathbf{C}$ on the elements of $\mathcal{S}$ can be defined by taking its action on the basis elements to be

$$(e_\alpha, e_\beta) = \mathrm{i}\omega(e_\alpha^*, e_\beta) = \frac{\mathrm{i}}{4\kappa_5^2} \int_{\Sigma_t} \mathrm{d}^3x\, \mathrm{d}y\, \frac{a^3}{n} \left(e_\alpha^* \dot{e}_\beta - e_\beta \dot{e}_\alpha^*\right). \tag{5.7.11}$$

One then extends the inner product (or $\omega$) to the whole of $\mathcal{S}$ (possibly complexified) by linearity. The space $\mathcal{S}$ with symplectic form $\omega$ can be identified with the phase space $(M, \omega)$ constructed in Section 2.1 (Wald, 1994). Each point $y \in M$ specifies that configuration variables and their canonical momenta, which if the theory possesses a well-defined initial

[24]The general issue of quantum field theory in non-globally hyperbolic spacetimes continues to be worked on. See, for example, Ishibashi and Wald (2004); Kay (1996).

value problem[25] can be used to construct a solution of the equations of motion with initial values at $t = 0$ coinciding with $y$. Therefore one may freely identify $M$ and $\mathcal{S}$.

By rewriting the inner product on $\mathcal{S}$ covariantly, one discovers

$$(e_\alpha, e_\beta) = \frac{\mathrm{i}}{4\kappa_5^2} \int_{\Sigma_t} \mathrm{d}\Sigma^a \left(e_\alpha^* \nabla_a e_\beta - e_\beta \nabla_a e_\alpha^*\right), \tag{5.7.12}$$

in which general coordinate invariance is obvious. The inner product therefore corresponds to the presence of a conserved current $j_a$,

$$j_a^{(\alpha)} = \frac{\mathrm{i}}{4\kappa_5^2} \left(e_\alpha^* \nabla_a e_\alpha - e_\alpha \nabla_a e_\alpha^*\right). \tag{5.7.13}$$

This is the Klein–Gordon current, as we anticipated. The construction of the symplectic form $\omega$ fixes the overall scale of $j_a$ from the action. The flux of particles across a slice $y =$ constant can be obtained by integration,

$$F = \frac{\mathrm{i}}{4\kappa_5^2} \int_{y=y_0} \mathrm{d}\Sigma^a (e_\alpha^* \nabla_a e_\alpha - e_\alpha \nabla_a e_\alpha^*), \tag{5.7.14}$$

which vanishes if $e_\alpha$ is real.[26] Clearly, any zero-mode of the conventional form $\partial e_\alpha / \partial y = 0$ will have no particle flux across any $y$-slice: one says that the zero mode is bound to the brane.

The remaining task is to enumerate the basis elements of the Hilbert space. In analogy with the general framework of string compactification, we look for separable solutions to the field equation, of the form

$$e_\alpha(\mathbf{k}) = (2\pi)^{-3} \mathcal{E}_\alpha(y) \varphi_\alpha(t; \mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}. \tag{5.7.16}$$

Let us specialise to the case of Kaloper–Linde branes, carrying an inflating de Sitter cosmology. The analogous theory of Randall–Sundrum gravitons can be found in the original

[25] In anti-de Sitter space there is again a problem, because anti-de Sitter space does not possess a well-defined initial value problem in the usual sense: there are no Cauchy surfaces (Hawking and Ellis, 1973). Nonetheless, as indicated above, these difficulties can be overcome. In this passage we are really describing the isomorphism between phase space and the manifold $\mathcal{S}$ of solutions in the globally hyperbolic case, which suffices to illustrate the general point.

[26] More strictly, the requirement is that conjugation introduces only a function of $t$ and $\mathbf{x}$,

$$e_\alpha^* = f(t, \mathbf{x}) e_\alpha. \tag{5.7.15}$$

Therefore $|f|^2 = 1$.

## 5.6. Stability of the brane

Having described the various major families of brane universe, the next step in formalizing a secure scaffold on which to build a cosmology of our universe is to establish their stability. We briefly alluded to problems with the stability of brane worlds when discussing the lack of anti-de Sitter no-hair theorem above. This means it is not possible to prove rigorous results which demonstrate, for example, the end-point in the evolution of all brane cosmologies which carry a non-zero cosmological constant is the exact de Sitter world. This may turn out to be of some considerable relevance to our universe, which appears on the basis of current observations to fall into this category. One might imagine that it is perfectly sensible to try and generalize Wald's four-dimensional argument to a cosmology carried by the brane, but the problem arises when one tries to constrain possible information falling onto the brane from far in the bulk. There is no way to show that bulk perturbations smooth out over time, so any de Sitter state in the asymptotic future could always be disrupted by new material falling onto the brane.

The no-hair theorem forms one half of a pair of important results which buttress our understanding of the vacua of general relativity in four dimensions. The other is the positive energy theorem of Schoen and Yau (1979). This extremely non-trivial result can be understood as a higher-dimensional generalization of the classical singularity theorems of Penrose, Hawking and Geroch (Hawking and Ellis, 1973). In proving these theorems, one argues by contradiction, constructing putative worldlines of hypothetical observers whose existence is contradictory, whereas Schoen and Yau constructed surfaces with similar properties. The proof can be considerably simplified using twistor techniques (Witten, 1981), and in this form can be fairly straightforwardly generalized to anti-de Sitter space (Flaherty, 1984; Perry, 2001) via a supercovariantization argument. There is no analogous extension of the result to de Sitter space. This happens for a good reason: de Sitter space is essentially unstable to the nucleation of black holes, because the path integral has a saddle at the Nariai metric, which is the instanton for black hole production immersed in a de Sitter background (Perry, 1982).

---

Taken together with the conservation equation $\dot{\rho} + 3H(p + \rho) = 0$, this is equivalent to

$$\dot{H} + H^2 = -\frac{\kappa_4^2}{6}\left[(\rho + 3p)\left(1 + \frac{\rho}{\lambda}\right) + \frac{\rho^2}{\lambda}\right]. \tag{5.5.33}$$

The left-hand side is occasionally written as $\ddot{a}/a$.

These results solidify our understanding of the maximally symmetric vacua of general relativity in immediately useful ways. The Schoen–Yau/Witten positive energy results guarantee that Minkowski space cannot decay into some other vacuum state, and the same applies to anti-de Sitter space. De Sitter space may nucleate black holes,[20] but is an attractor state for any universe with positive vacuum energy. Exactly the same concerns occur in the braneworld: one wishes to show that a universe carried on a Minkowski brane, for example, will not decay into some other state, or at least that such decays happen on a timescale much longer than the age of the universe. However, the stringy character of the compactification radically alters our expectation about how much it will be possible to rigorously prove. The current position can be conveniently summarised:

- There is no positive energy theorem even for Minkowski branes. One can solve the asymptotic twistor equation (Stewart, 1991; Witten, 1981) to reproduce Witten's spinorial argument from four dimensions, but the result for the mass at infinity is not positive definite.[21] This happens because the strong energy condition, which guarantees positivity in four dimensions, is not sufficiently strong to govern the behaviour of the quadratic tensor $\tau_{ab}$ which appears in the brane Einstein equations, and, moreover, the tensor $\mathsf{E}_{ab}$ is not constrained at all. (However for BPS brane configurations in a supergravity vacuum, such as the theory of supersymmetric Randall–Sundrum branes, one expects positivity results in the normal sense; see Bergshoeff, Kallosh, and Van Proeyen (2000, 2001).)
- The no-hair theorem can be generalized to the brane (Santos, Vernizzi, and Ferreira, 2001), although in slightly weakened form. In particular, any strong

[20]This may be of some importance during a vacuum dominated epoch of the Universe's evolution. However, the absence of rigorous stability results should be interpreted as a reflection of the peculiarity of de Sitter space within string theory (Kachru et al., 2003b) as a resonant state which must eventually decay into some other spacetime in which future null infinity is not spacelike.

[21]This calculation does not seem to have appeared in the literature. However it is a trivial modification of the argument from four dimensions. No use is made of the Einstein equations in Witten's argument until it is necessary to replace the curvature tensor $R_{ab}$ in terms of the energy and momentum carried by the space time. This is the crucial step which fails, because of the tensor $\mathsf{F}_{ab}$, which cannot be predicted on the basis of data given on the brane.

anisotropic stress may cause the brane to collapse. However, for sufficiently sanitised inflationary models the homogeneous, isotropic de Sitter world is still the end-point of inflationary evolution.

- The situation regarding black holes on the braneworld is presently undecided (Chamblin, Hawking, and Reall, 2000; Guedens, Clancy, and Liddle, 2002; Gutowski and Reall, 2004; Kudoh and Wiseman, 2004; Reall, 2003). In particular, an exact black hole metric on the brane is not yet known. However, there is no reason to believe that de Sitter space on the brane will not be unstable to black hole nucleation in the same way as in four dimensions. This is an issue of some considerable interest, and we will return to it briefly at the end of Chapter 9.

**5.6.1. Tachyon matter and brane instabilities.** The conclusion from the absence of rigorous stability results is that brane vacua may generically exhibit instabilities of one form or another. The reason for this is not hard to find, and arises from the observation that branes are dynamical objects in string theory. The exchange of string modes between branes may cause attractive or repulsive forces to change the configuration. This need not be a bad thing: the cyclic and Ekpyrotic scenarios, for example, rely on this very mechanism. In simple cases exchange of virtual string modes may only cause the various moduli fields which parametrize the vacuum, such as the inter-brane distance, to roll. In more complicated scenarios the vacuum itself may change as a phase transition occurs which causes a shift to some other vacuum. For example, new branes may condense out of the vacuum or topologically non-trivial field configurations may be generated by the collision of branes (Gray, 2004). Instabilities of this sort are rather generally signalled by the presence of tachyonic fields in the bulk. In particular, an interpretation of the open string tachyon of bosonic string theory strongly advocated by Sen (Sen, 1998, 1999, 2002a,b, 2003) characterizes the vacuum of bosonic string theory as an unstable background built out of branes. The bosonic string tachyon is a field mode which describes the propensity of the branes in this background to annihilate. (See also Frolov, Kofman, and Starobinsky (2002); Gibbons (2003).)

Exactly the same effect can be expected to occur in cosmological brane compactifications Frolov and Kofman (2004). In this case when one includes a bulk scalar field, this field typically displays tachyonic instabilities. There are various ways in which one can attempt to stabilize the vacuum (Goldberger and Wise, 2002), but appeal to the higher

dimensional string theory shows that in the absence of supersymmetry the vacuum is likely to be unstable. We do not address stability issues in this work, ascribing the problem to the general issue of constructing sane string compactifications.

## 5.7. BDEL brane compactifications and zero modes of the graviton

Since the basic brane world geometries, together with their attendant interpretations in terms of stringy physics (or CFTs), and some simple properties like stability have now been established, or at least clarified, the final step is to try and describe low energy physics from the point of view of an observer sitting on one of the branes. The principal low energy excitations are the Kaluza–Klein modes of the supergravity fields in the bulk. In a traditional string theory compactification these Kaluza–Klein modes would be described by the technology outlined in Section 3.2. The metric for the theory, containing gravitational excitations of all types, is

$$\mathrm{d}s^2 = -n^2(1+2\phi)\mathrm{d}t^2 + a^2(\delta_{ij}+e_{ij})\mathrm{d}x^i\mathrm{d}x^j + 2A_i\mathrm{d}x^i\mathrm{d}t + \mathrm{d}y^2 \tag{5.7.1}$$

which consists of a scalar $\phi$, a vector $A_i$, and a tensor $e_{ij}$. We have used some of the gauge freedom to retain Gaussian normal coordinates, meaning that there is no perturbation involving $y$-components of the metric. This is no more than the simple the Kaluza–Klein decomposition which we have already seen, Eq. (3.2.1). The Kaluza–Klein vector $A_i$ and scalar $\phi$ have no zero-modes in the sense of (3.2.5b), but carry a spectrum of massive excitations with mass $m > 3H/2$ (Bridgman, Malik, and Wands, 2002; Frolov and Kofman, 2002).[22] These excitations are not important for inflation, so we ignore $A_i$ and $\phi$ in what follows, concentrating only on the symmetric tensor piece $e_{ij}$.

We will frequently need the curvature quantities corresponding to this metric. For the purposes of this chapter, we are only interested in building the $e_{ij}$ field equation, so

[22] In talking about Kaluza–Klein modes we are being slightly sloppy with notation, since the Kaluza–Klein fields *per se* exist only in the case where the brane world can be given a bundle structure, as described above, or more physically where the wave equation is separable: in physical terms, this means the cosmology carried by the brane is maximally symmetric. This special case is assumed universally throughout the present section. Notice that it is for this reason that we expended so much effort in describing the topological stability of the zero mode when discussing brane world compactifications, because eventually it will be important to consider universes which are marginally perturbed away from maximal symmetry. In this case, the Kaluza–Klein decomposition will not exist and it is no longer clear that a normalizable zero mode persists.

a calculation to first order in $e_{ij}$ will suffice. In later chapters we will be working in the path integral formalism, where the appropriate quantity is not the field equation but the action: in this case, a more careful approach is needed, and one must evaluate the action to second order in the fields. This will be done explicitly in Chapter 6, but would be needlessly troublesome to obtain here. We take the matter content to be the same as (5.1.2), including a bulk source $Q_{ab}$ and a brane energy–momentum tensor $S_{ab}$. The equations governing $a$ and $n$ match the unperturbed case, whereas one can find an evolution equation for $e_{ij}$ by writing down the $(i,j)$ Einstein equation,

$$\frac{a^2}{2}\Box e_{ij} + (\delta_{ij} + e_{ij})Z_0 = \Lambda g_{ij} - \lambda\delta_{\mathrm{D}}(y)h_{ij} + \kappa_5^2\left[Q_{ij} - \delta_{\mathrm{D}}(y)S_{ij}\right], \tag{5.7.2}$$

where $\Box$ is the *scalar* background Laplacian,

$$\Box = \frac{1}{\sqrt{-g_0}}\partial_a\left(\sqrt{-g_0}g_0^{ab}\partial_b\right), \tag{5.7.3}$$

and $Z_0$ is a background geometrical quantity,

$$Z_0 = a'^2 - \frac{\dot{a}^2}{n^2} + 2aa'' - 2\frac{a\ddot{a}}{n^2} + 2aa'\frac{n'}{n} - 2\frac{a\dot{a}}{n^2}\frac{\dot{n}}{n} + a^2\frac{n''}{n}. \tag{5.7.4}$$

The value of $Z_0$ is set by the background Einstein equations, and is most easily found by picking out the $O(1)$ part of (5.7.2). We find

$$\frac{Z_0}{a^2} = \Lambda - \lambda\delta_{\mathrm{D}}(y), \tag{5.7.5}$$

which allows $Z_0$ to be eliminated from the $O(e)$ equation, to give

$$\frac{a^2}{2}\Box e_{ij} = \kappa_5^2(Q_{ij} + \delta_{\mathrm{D}}(y)S_{ij}). \tag{5.7.6}$$

Integrating over a small neighbourhood of $y = 0$ shows that $e_{ij}$ obeys the Israel jump condition (Langlois, Maartens, and Wands, 2000)

$$[e'_{ij}]^+_- = -\frac{2n^2}{a^2}S_{ij}. \tag{5.7.7}$$

If there is no anisotropic stress, then $[e'_{ij}]^+_- = 0$. This is a strong restriction on the kinds of boundary conditions which are allowed at the brane, for which tensor fields such as $e_{ij}$ are much more constrained than lower spin fields (Flachi, Moss, and Toms, 2001; Flachi and Toms, 2001).

If there are no sources $Q_{ab}$, $S_{ab}$ then the field equation is $\Box e_{ij} = 0$ and one can construct by hand an effective action,

$$S_g = \frac{1}{8\kappa_5^2} \int \mathrm{d}t\, \mathrm{d}^3x\, \mathrm{d}y\; na^2 \left( \frac{1}{n^2} \dot{e}_{ij} \dot{e}^{ij} - \frac{1}{a^2} \partial^k e_{ij} \partial_k e^{ij} - e'_{ij} e'^{ij} \right). \tag{5.7.8}$$

Up to possible boundary terms or cocycles, this Lagrangian describes the on-shell graviton.[23]

**5.7.1. Canonical quantization.** Let us quantize the field $e_{ij}$ using the canonical procedure. This is fairly awkward, and in any case will not be the preferred method when dealing with the more complicated situations encountered in Part 2. We give the canonical treatment here in order to provide a sound point of departure for the path integral formalism, and so that we can contrast the approaches.

The field canonically conjugate to $e_{ij}$ is

$$\pi^{ij} = \frac{\delta S_g}{\delta \dot{e}_{ij}} = \frac{1}{4\kappa_5^2} \frac{a^3}{n} \dot{e}^{ij}. \tag{5.7.9}$$

The field equation for $e_{ij}$ is the Klein–Gordon equation for the tensor $e_{ij}$, considered as a scalar on spacetime indices. Therefore the effective action must be the Klein–Gordon equation in unusual coordinates, and quantization could be effected by repeating the familiar process of Klein–Gordon quantization on curved backgrounds (Birrell and Davies, 1982; Wald, 1994). This would involve building the conserved Klein–Gordon current $j_a$, which arises from $U(1)$ phase transformations of the field $e_{ij}$, and converting $j_a$ to a conserved inner product. The quantization consists of building the quantum field $\hat{e}_{ij}$ out of the Hilbert space of postive frequency eigensolutions to the field equation equipped with the Klein–Gordon inner product. Fortunately, there is a constructive alternative which essentially consists of the argument of Section 2.1 in reverse (Carlip, 1998; Wald, 1994).

Let $e_\alpha$ be a family of solutions to the field equation $\Box e_\alpha = 0$, with the internal $\mathfrak{so}(3)$ indices omitted for clarity. One can form a symplectic product on these solutions by the rule

$$\omega(e_\alpha, e_\beta) = \int_{\Sigma_t} \mathrm{d}^3x\, \mathrm{d}y\; (e_\alpha \pi_\beta - e_\beta \pi_\alpha) = \frac{1}{4\kappa_5^2} \int_{\Sigma_t} \mathrm{d}^3x\, \mathrm{d}y\; \frac{a^3}{n} (e_\alpha \dot{e}_\beta - e_\beta \dot{e}_\alpha), \tag{5.7.10}$$

where the integration is taken over a spatial hypersurface $\Sigma_t$ (ordinarily a Cauchy surface, but in the anti-de Sitter case there are some subtleties: the ordinary theory of quantization

[23] It is useless when working off-shell, as one does when properly fixing the gauge, to be described in Chapter 6. The action has been artificially gauge-fixed by making $e_{ij}$ transverse and traceless.

on curved manifolds is only guaranteed to work in globally hyperbolic spacetimes (Wald, 1994). Anti-de Sitter space is not globally hyperbolic (Hawking and Ellis, 1973), so there is a difficulty. This problem was resolved and a satisfactory theory of quantization in anti-de Sitter space was worked out in the context of gauged extended supergravity (see Mezincescu and Townsend (1985), and Ceresole and Dall'Agata (2000) four the five-dimensional case relevant to the case at hand) in a series of papers completed in the mid-1980s (Avis, Isham, and Storey, 1978; Breitenlohner and Freedman, 1982a,b; Gibbons, Hull, and Warner, 1983; Mezincescu and Townsend, 1985).[24] The quantization prescription which was developed consists in compactifying AdS and imposing boundary conditions on the solutions to the classical equations of motion entering the Hilbert space of the quantum theory: namely, that the Hilbert space inner product to be constructed below actually is conserved, and that neither particles nor energy are exchanged with the boundary (Balasubramanian, Kraus, and Lawrence, 1999; Klebanov and Witten, 1999; Minces and Rivelles, 2001; Witten, 1998). This prescription was used to understand the Hilbert space of bulk fields in the AdS/CFT correspondence described in Section 5.4 above, and forms the basis for the division of bulk fields into normalizable and non-normalizable modes, the first of which fluctuate and enter the bulk Hilbert space, and the latter of which do not. This is important in AdS/CFT for understanding the appearance of four-dimensional gravity on the brane, as we have described.

Let $\mathcal{S}$ be the space of classical solutions to the equations of motion, and now suppose that the $e_\alpha$ span $\mathcal{S}$. An inner product $(\ ,\ ) : \mathcal{S} \times \mathcal{S} \to \mathbf{C}$ on the elements of $\mathcal{S}$ can be defined by taking its action on the basis elements to be

$$(e_\alpha, e_\beta) = \mathrm{i}\omega(e_\alpha^*, e_\beta) = \frac{\mathrm{i}}{4\kappa_5^2} \int_{\Sigma_t} \mathrm{d}^3x\, \mathrm{d}y\, \frac{a^3}{n} \left(e_\alpha^* \dot{e}_\beta - e_\beta \dot{e}_\alpha^*\right). \tag{5.7.11}$$

One then extends the inner product (or $\omega$) to the whole of $\mathcal{S}$ (possibly complexified) by linearity. The space $\mathcal{S}$ with symplectic form $\omega$ can be identified with the phase space $(M, \omega)$ constructed in Section 2.1 (Wald, 1994). Each point $y \in M$ specifies that configuration variables and their canonical momenta, which if the theory possesses a well-defined initial

[24] The general issue of quantum field theory in non-globally hyperbolic spacetimes continues to be worked on. See, for example, Ishibashi and Wald (2004); Kay (1996).

value problem[25] can be used to construct a solution of the equations of motion with initial values at $t = 0$ coinciding with $y$. Therefore one may freely identify $M$ and $\mathcal{S}$.

By rewriting the inner product on $\mathcal{S}$ covariantly, one discovers

$$(e_\alpha, e_\beta) = \frac{\mathrm{i}}{4\kappa_5^2} \int_{\Sigma_t} \mathrm{d}\Sigma^a \, (e_\alpha^* \nabla_a e_\beta - e_\beta \nabla_a e_\alpha^*) \,, \tag{5.7.12}$$

in which general coordinate invariance is obvious. The inner product therefore corresponds to the presence of a conserved current $j_a$,

$$j_a^{(\alpha)} = \frac{\mathrm{i}}{4\kappa_5^2} \, (e_\alpha^* \nabla_a e_\alpha - e_\alpha \nabla_a e_\alpha^*) \,. \tag{5.7.13}$$

This is the Klein–Gordon current, as we anticipated. The construction of the symplectic form $\omega$ fixes the overall scale of $j_a$ from the action. The flux of particles across a slice $y =$ constant can be obtained by integration,

$$F = \frac{\mathrm{i}}{4\kappa_5^2} \int_{y=y_0} \mathrm{d}\Sigma^a (e_\alpha^* \nabla_a e_\alpha - e_\alpha \nabla_a e_\alpha^*), \tag{5.7.14}$$

which vanishes if $e_\alpha$ is real.[26] Clearly, any zero-mode of the conventional form $\partial e_\alpha / \partial y = 0$ will have no particle flux across any $y$-slice: one says that the zero mode is bound to the brane.

The remaining task is to enumerate the basis elements of the Hilbert space. In analogy with the general framework of string compactification, we look for separable solutions to the field equation, of the form

$$e_\alpha(\mathbf{k}) = (2\pi)^{-3} \mathcal{E}_\alpha(y) \varphi_\alpha(t; \mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}. \tag{5.7.16}$$

Let us specialise to the case of Kaloper–Linde branes, carrying an inflating de Sitter cosmology. The analogous theory of Randall–Sundrum gravitons can be found in the original

[25] In anti-de Sitter space there is again a problem, because anti-de Sitter space does not possess a well-defined initial value problem in the usual sense: there are no Cauchy surfaces (Hawking and Ellis, 1973). Nonetheless, as indicated above, these difficulties can be overcome. In this passage we are really describing the isomorphism between phase space and the manifold $\mathcal{S}$ of solutions in the globally hyperbolic case, which suffices to illustrate the general point.

[26] More strictly, the requirement is that conjugation introduces only a function of $t$ and $\mathbf{x}$,

$$e_\alpha^* = f(t, \mathbf{x}) e_\alpha. \tag{5.7.15}$$

Therefore $|f|^2 = 1$.

paper, Randall and Sundrum (1999b), or Giudice, Kolb, Lesgourgues, and Riotto (2002). The inner product on the modes $e_\alpha$ is

$$\begin{aligned}(e_\alpha(\mathbf{k}), e_\beta(\mathbf{k}')) &= \frac{\mathrm{i}}{4\kappa_5^2}\int \mathrm{d}^3x\,\mathrm{d}y\;\mathcal{A}^2(y)a_\mathrm{b}^3(t)[e_\alpha^*\dot{e}_\beta - e_\beta\dot{e}_\alpha^*] \\ &= \frac{\mathrm{i}}{4\kappa_5^2}\frac{a_\mathrm{b}^3}{(2\pi)^3}\triangle[\varphi_\alpha^*(\mathbf{k}), \varphi_\beta(\mathbf{k}')]\delta_\mathrm{D}(\mathbf{k}-\mathbf{k}')\int \mathrm{d}y\;\mathcal{A}^2\mathcal{E}_\alpha^*\mathcal{E}_\beta,\end{aligned} \tag{5.7.17}$$

where $\triangle[f,g]$ is the $t$-Wronskian

$$\triangle[f,g] = f\dot{g} - \dot{f}g. \tag{5.7.18}$$

Imposing the field equation $\Box e_\alpha$ on (5.7.16) gives two separate field equations for $\mathcal{E}$ and $\varphi$,

$$\ddot{\varphi}_\alpha + 3H\varphi_\alpha + \left(\alpha^2 + \frac{k^2}{a_\mathrm{b}^2}\right)\varphi_\alpha = 0 \tag{5.7.19a}$$

$$\mathcal{E}_\alpha'' + 4\frac{\mathcal{A}'}{\mathcal{A}}\mathcal{E}_\alpha' + \frac{\alpha^2}{\mathcal{A}^2}\mathcal{E} = 0. \tag{5.7.19b}$$

To understand how these equations relate to the Kaluza–Klein equations (3.2.5a)–(3.2.5b) for a generic string compactification, it is necessary to realize that the Randall–Sundrum and Kaloper–Linde models represent a generalization of the direct product compactifications $M_4 \times K_6$ which were discussed in Section 3.2.1. Models such as these (and a whole class of $Dp$-brane metrics) have a rather more general line element[27]

$$\mathrm{d}s_{10}^2 = \mathrm{e}^{-2\alpha(y)}\mathrm{d}s_4^2 + \mathrm{e}^{2\alpha(y)}\mathrm{d}s_6^2. \tag{5.7.20}$$

This is not a direct product. Instead, one sometimes refers to compactifications of the form (5.7.20) as warped compactifications. The four-dimensional metric $\mathrm{d}s_4^2$ is assumed to be independent of $y$. As in Section 3.2.1, the 6-manifold $K_6$ is here trimmed to a single extra dimension, in accordance with the general principles outlined before Eq. (3.4.5), and $\mathrm{e}^{2\alpha(y)}\mathrm{d}s_6^2 = \mathrm{d}y^2$. Such metrics are best understood as fibre bundles (cf. Section B.4 (Motl, 2003). In the Kaloper–Linde case (for example), with metric

$$\mathrm{d}s^2 = \mathcal{A}^2\left(-\mathrm{d}t^2 + a_\mathrm{b}^2\mathrm{d}\mathbf{x}^2\right) + \mathrm{d}y^2, \tag{5.7.21}$$

one can interpret the full spacetime as a fibration of our universe, with metric

$$\mathrm{d}s_4^2 = -\mathrm{d}t^2 + a_\mathrm{b}^2\mathrm{d}\mathbf{x}^2, \tag{5.7.22}$$

[27]This section is partially based on lectures given by Hermann Verlinde at the Insitute for Advanced Study, Verlinde (2003).

over a circle. The Laplacian on this space takes the form[28]

$$\begin{aligned}\Box &= \frac{1}{\mathcal{A}^2\sqrt{-g_4}}\partial_i(\mathcal{A}^4\sqrt{-g_4}g_4^{ij}\partial_j) + \frac{1}{\mathcal{A}^4}\partial_y(\mathcal{A}^4\partial_y) \\ &= \Box_4 + \frac{1}{\mathcal{A}^4}\partial_y(\mathcal{A}^4\partial_y)\end{aligned} \tag{5.7.25}$$

This consists of the fibre Laplacian, $\Box_4$, which is the Laplacian on $\mathrm{d}s^4$, together with a piece related to the the base space Laplacian. This is rather different to the direct product case, Eq. (3.2.5a). By replacing the wavefunctions $\psi$ on which $\Box$ operates with a mangled version $\psi_0$, defined by $\psi = \mathcal{A}^{-2}\psi_0$, we recover

$$\Box\psi = \mathcal{A}^{-4}\Box_4\psi_0 + \partial_y^2\psi_0 - 2(\frac{\mathcal{A}''}{\mathcal{A}} + \frac{\mathcal{A}'}{\mathcal{A}}\frac{\mathcal{A}'}{\mathcal{A}})\psi_0. \tag{5.7.26}$$

This has the form

$$\begin{aligned}&\left[\mathcal{A}^{-4}\,(\text{fibre Laplacian}) + \text{mass piece} + (\text{base Laplacian})\right]\psi_0 = 0 \\ \text{or, equivalently}\quad &\left[\mathcal{A}^{-4}\,(\text{fibre Laplacian}) + (\text{mangled base Laplacian})\right]\psi_0 = 0.\end{aligned} \tag{5.7.27}$$

where the mangled base Laplacian is the base Laplacian plus a transverse-dependent mass term,

$$\text{mangled base Laplacian} = \partial^2 - 2(\frac{\mathcal{A}''}{\mathcal{A}} + \frac{\mathcal{A}'}{\mathcal{A}}\frac{\mathcal{A}'}{\mathcal{A}}). \tag{5.7.28}$$

One is now back to the direct product case $M_4 \times K_6$, where the full Laplacian $\Box$ decomposes into a sum of Laplacians on the factors, except that instead of dealing with the base Laplacian, one should instead use the mangled Laplacian. All the rules enumerated in Section 3.2.1 concerning zero modes, and in particular the relation between the number of such modes and the topological Betti number $b_0$ of the compactification manifold, can now be carried over wholesale.

---

[28]For comparison, the two-dimensional flat Euclidean plane written in polar coordinates is another very simple example of a warped compactification, which is more familiar than the braneworld metric. This metric in this case is

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2. \tag{5.7.23}$$

The Laplacian is

$$\Box = \frac{1}{\sqrt{g}}\partial_a(\sqrt{g}g^{ab}\partial_b) = \frac{1}{r}\partial_r(r\partial_r) + \frac{1}{r^2}\partial_\theta^2 = \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\theta^2. \tag{5.7.24}$$

In this case, one interprets flat $\mathbf{R}^2$ as a fibre bundle with standard fibre $\mathbf{S}^1$ and base space the radial semi-infinite line $\mathbf{R}^+$, corresponding to the coordinate $r$. This contains the Laplacian with respect to the base space $(\partial^2/\partial r^2)$ together with the Laplacian on the fibre $(\partial^2/\partial\theta^2)$, multiplied by a function of the warp factor $(r^{-2})$, plus some other pieces, as described in the text.

- Randall–Sundrum model. The warp factor is $\mathcal{A} = \mathrm{e}^{-y\ell^{-1}}$, so the mangled Laplacian is

$$\partial_y^2 - \frac{4}{\ell^2}, \tag{5.7.29}$$

and the AdS curvature scale acts as a direct mass shift.

- Kaloper–Linde model. Now the warp factor is $\mathcal{A} = (H\ell/\sqrt{2})(\cosh 2\ell^{-1}(y_\mathrm{h} - y) - 1)^{1/2}$. The mangled Laplacian is

$$\partial_y^2 - \frac{4}{\ell^2}\left(1 - \frac{1}{\cosh 2\ell^{-1}(y_\mathrm{h} - y)}\right), \tag{5.7.30}$$

in which the plain mass shift is joined by a $y$-dependent piece. In view of the connexion between the $y$-direction and the renormalization group discussed above, this term is not unexpected.

In the one-dimensional case (which is essentially topologically trivial anyway, the only significant choice being the between compact circle $\mathbf{S}$ on the non-compact real line $\mathbf{R}$) there is a zero mode corresponding to $\psi_0 = \mathcal{A}^2$. As in the case of string compactification, this zero mode is *stable* under perturbations of the standard fibre – it is a toplogical property of the compactification manifold (see also Chamblin and Gibbons (2000)).

Returning to the (effectively) mangled compactification Laplacian (5.7.19b), this can be cast in Sturm–Liouville form (cf. (A.1.1)), viz.

$$(\mathcal{A}^4\mathcal{E}')' + \alpha^2\mathcal{A}^2\mathcal{E} = \mathcal{L}\mathcal{E} + m^2\mathcal{A}^2\mathcal{E} = 0, \tag{5.7.31}$$

where $\mathcal{L}$ is the corresponding Liouville operator. This is a perfectly well-posed Sturm–Liouville problem, for which $\mathcal{L}$ is self-adjoint provided

$$\left[\mathcal{E}_\alpha^*\mathcal{A}^4\mathcal{E}_\beta' - \mathcal{E}_\beta\mathcal{A}^4\mathcal{E}_\alpha^{*\prime}\right]_{-y_\mathrm{h}}^{+y_\mathrm{h}} = 0. \tag{5.7.32}$$

For $\mathcal{E}_\alpha$ obeying this condition, we can choose the convenional normalization (via the argument leading to (A.1.4))

$$\int \mathrm{d}y\ \mathcal{A}^2\mathcal{E}_\alpha^*\mathcal{E}_\beta = \delta_{\alpha\beta}. \tag{5.7.33}$$

The rest of the normalization rule (5.7.17) then amounts to a Wronskian condition on $\varphi_\alpha$,

$$\triangle[\varphi(\mathbf{k}), \varphi(\mathbf{k}')] = -4\mathrm{i}\kappa_5^2(2\pi)^3 a_\mathrm{b}^{-3}. \tag{5.7.34}$$

This normalization on the fibre modes $\varphi_\alpha$ is just the Klein–Gordon normalization which is familiar from the study of four-dimensional inflation.

At this point the conceptual elements of the construction are complete, and one is only left with the task of solving the mangled Laplacian (or (5.7.19b)) to find its eigenvalue spectrum. These eigenvalues, as we have seen, correspond to the masses of particles in four dimensions. Since the details of the computation are purely mechanical, we simply refer to the literature (Gorbunov, Rubakov, and Sibiryakov, 2001) for the technical aspects. There is an equivalent, simpler way to recover the spectrum which we present here.

Let $z = y_{\rm h} - y$, so that $z \to 0$ is the limit in which one approaches the horizon bounding the causal patch which surrounds the brane. We aim to calculate the behaviour of the $\mathcal{E}_\alpha$ as $z \to 0$. In this régime, the scale factor $\mathcal{A}(z)$ is approximately

$$\mathcal{A}(z) = H\ell \sinh z\ell^{-1} \sim Hz, \tag{5.7.35}$$

so the coefficients in (5.7.19b) behave like

$$\frac{\alpha^2}{\mathcal{A}^2} \to \frac{\alpha^2}{H^2 z^2} \quad \text{and} \quad 4\frac{\alpha'}{\alpha} \to -\frac{4}{z}. \tag{5.7.36}$$

Near $z = 0$, therefore, the behaviour of the $\mathcal{E}_\alpha$ is determined by the equation

$$\mathcal{E}_\alpha'' + \frac{4}{z}\mathcal{E}_\alpha' + \frac{\alpha^2}{H^2 z^2}\mathcal{E}_\alpha = 0. \tag{5.7.37}$$

This equation is equidimensional, with solutions of the form $\mathcal{E}_\alpha = \text{constant} \times z^{\Delta}(\alpha)$, where $\Delta(\alpha)$ is a scaling dimension determined by the rule

$$\Delta(\Delta + 3) = -\frac{m^2}{H^2} \quad \text{and therefore} \quad \Delta_\pm(\alpha) = -\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}. \tag{5.7.38}$$

In order to have oscillating modes (see Chapter A), one must have

$$m > \frac{3H}{2}. \tag{5.7.39}$$

In addition, there is the special case of the topological zero mode, as described above. There is no quantization condition on the allowed $m$. This is indicative of the spectrum of a non-compact differential operator, and arises in this case because one is encoding the details of a non-trivial conformal field theory (that is, not Gaussian) on the brane (Witten, 1999a).

The appearance of the scaling dimension (5.7.38) is not an accident (compare Eqs. (D.2.3)–(D.2.4)). It is in fact a completely standard relation in AdS/CFT for a field of mass $m = \alpha$ (with $m^2 \mapsto -m^2$ as appropriate in anti-de Sitter space; Balasubramanian et al. (1999); Klebanov and Witten (1999); Minces and Rivelles (2001); Witten (1998)) and dS/CFT

(Halyo, 2002a; Spradlin, Strominger, and Volovich, 2001; Strominger, 2001). In this context, $\Delta(m)$ describes the anomalous dimension $\gamma_{\mathcal{O}}$ of the CFT operator which couples to the bulk field $\phi$, with boundary behaviour $\phi \sim \phi_0 z^{\Delta}(m)$ (cf. (5.4.5)).[29] On the other hand, our interpretation of (5.7.38) is distinct from the usual AdS/CFT interpretation, which stems from Breitenlohner and Freedman (1982a,b), where the square root is required to be positive. If one instead makes this requirement, one does not recover the standard spectrum of KK modes. Instead, there is a positive dimension $\Delta_+$ which corresponds to an infinite-energy deformation of the background geometry, corresponding in AdS/CFT to a deformation of the dual field theory. The negative choice $\Delta_-$ leads to finite-energy perturbations which are naturally intepreted as fluctuations or excitations in the theory described by the dual Lagrangian.

**5.7.2. The zero mode.** Now let us specialise to the massless zero mode. The relevant harmonic form on the transverse compactification space is just a constant $\mathcal{E}_0 = \text{constant}$, which is normalized according to the Liouville rule

$$\int_K \mathrm{d}y\, \mathcal{A}^2 \mathcal{E}_0^2 = 1. \tag{5.7.40}$$

The corresponding forms on the de Sitter slices obey the Kaluza–Klein field equation

$$\ddot{\varphi} + 3H\dot{\varphi} + \frac{k^2}{a_{\mathrm{b}}^2}\varphi = 0, \tag{5.7.41}$$

and which are normalized according to (5.7.17),

$$\frac{\mathrm{i}}{4\kappa_5^2}\frac{a_{\mathrm{b}}^3}{(2\pi)^3}\triangle[\varphi^*(\mathbf{k}), \varphi(\mathbf{k})] = 1. \tag{5.7.42}$$

Let us change variables in the $\varphi$ equation by setting $u = a_{\mathrm{b}}\varphi$ and switching to conformal time $\tau$, defined by $\mathrm{d}\tau = a_{\mathrm{b}}^{-1}\,\mathrm{d}t$, after which the equation of motion reads

$$u'' + \left(k^2 - \frac{2}{\tau^2}\right) u = 0, \tag{5.7.43}$$

in which a prime $'$ denotes a $\tau$ derivative. The field $u$ should be normalized via

$$\triangle[u^*, u] = -\mathrm{i}\kappa_5^2(2\pi)^3. \tag{5.7.44}$$

with $\triangle$ now a $\tau$-Wronskian. The solution of the field equation is

$$u = A\sqrt{-k\tau}H_{3/2}^{(1)}(-k\tau) + B\sqrt{-k\tau}H_{3/2}^{(2)}(-k\tau), \tag{5.7.45}$$

[29] The holographic description of inflation is outlined in Appendix D.

where $A$ and $B$ are arbitrary complex coefficients. A straightforward calculation shows that the Wronskian of $u$ is

$$\triangle[u^*, u] = -k^2\tau\left(|A|^2 - |B|^2\right)\triangle[H^{(1)}_{3/2}, H^{(2)}_{3/2}], \tag{5.7.46}$$

where we have suppressed the $-k\tau$ dependence of the Hankel functions. It is a standard result (Morse and Feshbach, 1953) that as functions of $z$, the Wronskian of the Hankel functions satisfies

$$\triangle[H^{(1)}_\nu(z), H^{(2)}_\nu(z)] = \frac{4}{\pi i z}, \tag{5.7.47}$$

so substituting this into (5.7.44) gives

$$|A|^2 - |B|^2 = \frac{\pi\kappa_5^2}{k}(2\pi)^3. \tag{5.7.48}$$

For convenience, one can define complex numbers $c_1$, $c_2$, which can be thought of as parametrizing the vacuum state, by

$$A = c_1\kappa_5^2\sqrt{\frac{\pi(2\pi)^3}{k}} \quad \text{and} \quad B = c_2\kappa_5^2\sqrt{\frac{\pi(2\pi)^3}{k}}, \tag{5.7.49}$$

which then satisfy the vacuum condition $|c_1|^2 - |c_2|^2 = 1$ (Hwang, 1995). The field $\varphi$ has the general solution

$$\varphi = \frac{c_1\kappa_5}{a_{\rm b}}\sqrt{\frac{\pi(2\pi)^3}{k}}\sqrt{-k\tau}H^{(1)}_{3/2}(-k\tau) + \frac{c_2\kappa_5}{a_{\rm b}}\sqrt{\frac{\pi(2\pi)^3}{k}}\sqrt{-k\tau}H^{(2)}_{3/2}(-k\tau). \tag{5.7.50}$$

The Hilbert space consists of fields of this form, labelled by the various allowable $\mathbf{k}$-states. The quantum Kaluza–Klein zero mode field $\hat{e}$ is constructed by superposing an admixture of the $\mathbf{k}$-modes, with coefficients $a$, $a^\dagger$,

$$\hat{e} = \kappa_5\pi^{1/2}a_{\rm b}^{-1}\int\frac{{\rm d}^3k}{(2\pi)^{3/2}}\sqrt{\frac{-\tau}{k}}\mathcal{E}_0\left[a(\mathbf{k})(c_1H^{(1)}_{3/2} + c_2H^{(2)}_{3/2}){\rm e}^{{\rm i}\mathbf{k}\cdot\mathbf{x}} + a^\dagger(\mathbf{k})(c_1^*H^{(2)}_{3/2} + c_2^*H^{(1)}_{3/2}){\rm e}^{-{\rm i}\mathbf{k}\cdot\mathbf{x}}\right]. \tag{5.7.51}$$

In order that $\hat{e}$ and its canonical conjugate ($\hat{\pi}$ in our earlier notation) obey the canonical conjugation relations, the mode operators $a$ and $a^\dagger$ must commute according to the rules

$$[a(\mathbf{k}), a^\dagger(\mathbf{k}')] = \delta_{\rm D}(\mathbf{k} - \mathbf{k}'). \tag{5.7.52}$$

Apart from the prefactor $\mathcal{E}_0$, the standard calculation leading to the inflationary power spectrum (4.6.89) now applies, so

$$\Delta_e^2 = \frac{2}{\pi^2}\kappa_4^2F^2H^2|c_1 - c_2|^2, \tag{5.7.53}$$

where $\mathcal{E}_0^2 = \mu F^2$ is a conventional definition (Huey and Lidsey, 2001). To complete the picture one needs an expression for $F$, which is obtained from the normalization rule, $\mu F^2 = (2\int_0^{y_\mathrm{h}} \mathrm{d}y\, \mathcal{A}^2)^{-1}$. The integral is

$$\int_0^{y_\mathrm{h}} \mathrm{d}y\, \mathcal{A}^2 = \frac{x^2}{2}\int_0^{y_\mathrm{h}} (\cosh 2\ell^{-1}z - 1)\,\mathrm{d}z = \frac{x^2}{2}\left(\ell \sinh \ell^{-1}y_\mathrm{h} \cosh \ell^{-1}y_\mathrm{h} - y_\mathrm{h}\right), \qquad (5.7.54)$$

where we have introduced the common shorthand $x = H\ell$. This can be recast in the alternative form

$$F = (\mu\ell)^{-1/2}\left(\sqrt{1+x^2} - x^2 \sinh^{-1} x^{-1}\right)^{-1/2}, \qquad (5.7.55)$$

which matches with the expression given in Langlois et al. (2000) provided the Bunch–Davies vacuum $c_2 = 0$ is chosen and (as assumed by Langlois et al. (2000)) the four-dimensional cosmological constant $\Lambda_4$ vanishes, in which case $\ell = \mu^{-1}$.

**5.7.3. Path integral quantization.** The canonical method outlined in the previous section is lengthy, and necessitates the frustrating intermediate step of building a Hilbert space with conserved inner product. Although when one is seeking rigorous results this laborious formalism is unfortunately essential, it can profitably be skipped when one is dealing only with heuristic or intuitive arguments. For phenomenology, this is quite appropriate, and the path integral formalism can be applied instead. Moreover it is this technology which will generalize most easily to the more sophisticated examples of Part 2. This is a standard Kaluza–Klein analysis.

One begins with the effective action (5.7.8). After integrating by parts and discarding a surface term, we obtain

$$S_g = \frac{1}{8\kappa_5^2}\int \mathrm{d}t\,\mathrm{d}^3x\,\mathrm{d}y\,\left[\mathcal{A}^3 a_\mathrm{b}^3\left(\dot{e}_{ij}\dot{e}^{ij} - \frac{1}{a_\mathrm{b}^2}\partial^k e_{ij}\partial_k e^{ij}\right) + e_{ij}(\mathcal{A}^4 a_\mathrm{b}^3 e'^{ij})'\right]. \qquad (5.7.56)$$

By introducing eigenfunctions of the base Laplacian,

$$(\mathcal{A}^2\mathcal{E}_\alpha')' = -\alpha^2\mathcal{A}^2\mathcal{E}_\alpha \qquad (5.7.57)$$

in analogy with the decomposition of the equation of motion, the rules of Chapter A allow us to expand the field $e_{ij}$ as a superposition of the $\mathcal{E}_\alpha$,[30]

$$e_{ij} = \sum_\alpha f_{ij}^{(\alpha)}(t, \mathbf{x})\mathcal{E}_\alpha(y). \tag{5.7.58}$$

Using orthonormality and substituting into the action gives

$$S_g = \frac{1}{8\kappa_5^2}\sum_\alpha \int \mathrm{d}\tau\, \mathrm{d}^3x\, a_\mathrm{b}^2 \left( \dot{f}_{ij}^{(\alpha)}\dot{f}^{(\alpha)ij} - \partial^k f_{ij}^{(\alpha)}\partial_k f^{(\alpha)ij} - \frac{\alpha^2}{H^2\tau^2} f_{ij}^{(\alpha)} f^{(\alpha)ij} \right). \tag{5.7.59}$$

All trace of the transverse dimension has disappeared, except for the tower of Kaluza–Klein fields. In the path integral, the two-point function becomes (schematically)

$$\langle e(x)e(y)\rangle = \int [\mathrm{d}f]\mathcal{E}_0^2 f(x)f(y) \exp\left(\frac{\mathrm{i}}{\hbar}S_g[f]\right), \tag{5.7.60}$$

where we have discarded contributions except from the zero mode. Except for the renormalization $\mathcal{E}_0^2$, this is the path integral for a massless field in de Sitter space, and so reproduces the power spectrum found by operator methods.[31]

[30] If the theory is defined by Euclidean continuation, so that $Z = \int [\mathrm{d}e] \exp(-S_E)$ where $S_E$ is the Euclidean action, then any non-normalizable harmonics of the base manifold will not contribute to the path integral, since they will be exponentially supressed to zero. Therefore there is no need to impose any superselection rule which picks out the normalizable modes as the proper fluctuating wavefunctions which contribute to the quantum Hilbert space.

[31] We will carry out this calculation in detail in Chapter 7, so we avoid going into details here.

# Part 2

# Quantum braneworld phenomenology

CHAPTER 6

# Radiative constraints on brane quintessence

The cosmology outlined in preceding chapters has many attractive features, of which the application of ideas in string theory to cosmological questions is certainly not least. Global questions on cosmological scales can certainly provide a profitable testing ground for modern ideas in high energy physics, but although we can use the success or otherwise of cosmological models to inform our opinions concerning new ideas, the models themselves cannot be judged on these grounds and will ultimately stand or fall based on their phenomenological success, and, to some degree, their ability to reproduce the standard model.

By any measure, the dominant discovery of recent times, in cosmology and particle physics equally, is the observation that the expansion of the universe is accelerating, and not slowing down. This contradicts not only naïve expectation, but also well-founded cosmological theory based on the known gravitational properties of matter. The result is unambiguous, and confirmed by many experiments: the universe must largely be filled with a kind of matter quite unlike the baryonic matter with which we are familiar. This observation and its dependent consequences have illuminated and partly clarified the long-standing goal of carrying out a census of the global matter distribution in the universe, but a detailed microphysical explanation of the phenomenon itself, of the kind ordinarily provided by particle physics, is lacking. Although string-inspired cosmologies (Kane et al., 2003; Susskind, 2003) have some potential to provide an explanation, a convincing mechanism has yet to be found and it seems fair to say that progress in the short term will come from phenomenological or empirical approaches.

At its simplest, acceleration might be explained with nothing more than Einstein's abandoned cosmological constant. Although Einstein's reasons for both introducing and discarding $\Lambda$ have proved unreliable, in the modern context there is no reason to believe that $\Lambda$ should be absent. It is now understood that any integrated zero-point particle physics energy should contribute to $\Lambda$. On the other hand, naïve field theory calculations

based on this observation have been characterised as the worst estimate in the history of physics: estimates obtained in this way typically differ from observation by 120 orders of magnitude. It is possible to improve the situation somewhat by invoking supersymmetry, but generically it proves quite hard to construct supergravity vacua with positive cosmological constant (Townsend, 2001). The string theory case is even harder (Kachru et al., 2003b). Equally, if one takes the Einstein equations to appear either as an effective field theory approximating whatever quantum string or membrane theory encodes the true behaviour of gravity at short distances, or possibly to arise (as in string theory) as renormalization group equations enforcing the consistency of a background geometry, then $\Lambda$ will generically be present. In the absence of any theoretical control over $\Lambda$ itself, there is a strong temptation to explain the observations by invoking some other mechanism.

This chapter is concerned with some aspects of braneworld phenomenology. In particular, we shall examine some particle physics processes where one could reasonably expect the new mechanisms and points of view introduced by brane physics to provide substantial modifications to the conventional picture. The examples are drawn from proposals to help explain the observed acceleration, where the constraints and restrictions imposed by the standard model are particularly troubling. There is some possibility that braneworlds can alleviate such concerns, although the details are delicate and depend to some extent on factors which can be quite model specific.

In Section 6.1 we briefly discuss the background problem in four dimensions, and partially trace the history of proposals to explain the acceleration. In Section 6.2 we pass to the brane world, and translate some specific proposals to the new paradigm. This is all rather conventional, and can be found in numerous review articles and textbooks, so the presentation is short. To carry out detailed calculations, one needs a theory of particle physics on the brane world. A great deal of the standard formalism can be carried over verbatim, except for trivial changes in notation or dimensionality. However a new framework of Feynman diagrams must be provided which takes the place of the conventional field theory of weak-field Einstein gravity where scattering processes or interactions are mediated in an essential way by gravitons. We outline how this is accomplished in Section 6.6. It is then possible to study some representative processes in detail. We compute the corrections to the classical tree-level potential from a class of field-theory diagrams, and show that quite generally such corrections can be absorbed into an overall renormalization

which does not destroy the features of the potential. (On the other hand, we note that some couplings will be generated on curved manifolds which do not fall into this class of corrections, and which can generically be expected to non-trivially modify the potential.) Finally we compute a diagram contributing to the vacuum polarization of quintessence, and calculate an assessment of the order-of-magnitude of the mass shift which it generates. We show that this is generically rather small, so scalar fields falling into the quintessence class (including the inflaton) are safe from large Kaluza–Klein mediated radiative corrections of gravitational strength.

## 6.1. Quintessence and dark energy in four dimensions

No matter what exotic particle or type of matter eventually proves to be the correct explanation for acceleration, it can be said with some degree of certainty on the basis of many consistent experiments that the equation of state for such matter must be quite strange: it must support a negative pressure. This exotic behaviour, coupled with an evident non-luminous character, has resulted in this material being dubbed dark energy. Although in most respects this is a poor name, it has proved popular in the short term and most of the literature adopts this moniker, to which convention we shall conform. In fact there is no reason to believe that dark energy is any more like diffuse energy than conventional particulate matter, and it is quite possible that no new physical tools are required for its description, but rather the simple, conservative introduction of another quantum field to represent its particles. The nature of this quantum field remains mysterious. There have been proposals that work with rather high order fields, particularly 4-forms, but most workers exploit the obvious analogy with an early inflationary period and postulate the existence of a light scalar field moving under the influence of a potential. The scalar field is conventionally written $Q$ and called quintessence, since if $Q$ is massless or very light then it might mediate a sort of fifth force, of Yukawa form, with strength given by a modified inverse square law,

$$F \sim \frac{m\mathrm{e}^{-m_Q r}}{r^2}. \tag{6.1.1}$$

In the late 1960s and early 1970s superficial results concerning the oblateness of the sun, later contradicted, made a scalar–tensor theory of gravitation of Jordan type briefly popular. This was the Brans–Dicke theory. Although the original motivations for Brans–Dicke–Jordan theories have proved unfounded, there is widespread expectation, based on

the presence in closed string theory of a scalar dilaton field $\phi$ which would be part of the gravitational multiplet, that scalar–tensor gravitation may turn out to play a role in the effective description of nature at low energies after all. Experimental data concerning departures from pure general relativity within the solar system, originally gathered with an eye on Brans–Dicke gravity, provides very strong constraints on the mass and couplings of $Q$, and the range of any force which it might mediate, which can be carried over fairly straightforwardly to theories of quintessence. There are also proposals in which the kinetic rather than potential energy of the scalar field is dominant. Such theories are usually known as $k$-essence, and have improved tracking properties which can alleviate coincidence or 'why now' concerns (see below), but we will not address these models in any detail here.

The field $Q$ is a good example of an alternative mechanism by which an effective $\Lambda$ may arise, while keeping Einstein's faceless cosmological constant itself zero by some stringy or M-theory mechanism which at present escapes us. In the simplest models, the $Q$-field ought to have two quite distinct properties. On the one hand, it should not cluster on scales much smaller than the Hubble scale $H^{-1}$, which is roughly the same as the current horizon. On the other hand, it must be extremely light. In order not to contradict the successes of the standard cosmological picture, $Q$ must have induced acceleration only at very recent redshifts. It is possible to construct models in which $Q$ does not cluster, but is heavy, and vice-versa (Bassett, Kunz, Parkinson, and Ungarelli, 2003; Csaki, Kaloper, and Terning, 2002), but both now have observational support. In particular, the recent detection of cross-correlation between the Wilkinson Microwave Anisotropy Probe (WMAP) cosmic microwave background experiment and various tracers of large-scale structure (Boughn and Crittenden, 2003; Fosalba, Gaztanaga, and Castander, 2003; Scranton et al., 2003; Spergel et al., 2003), which are consistent with the decay of perturbations on large scales owing to an accelerating background, make construction of convincing non-accelerating models difficult.

Each of these properties gives rise to genuine concerns. Firstly, unless the potential $V(Q)$ is quite extraordinarily flat, $Q$ must have a large expectation value in order to be rolling sufficiently slowly that the slow-roll conditions are satisfied. A popular example in early universe inflation is known as chaotic inflation, in which the potential is assumed to be a polynomial $V(\phi) = \beta\phi^n/n!$ and one supposes that the universe is in some sense sufficiently

big that it is likely that some regions lie high enough on the potential for inflation to occur. However, this means that the inflaton field $\phi$ must acquire a super-Planckian expectation value, so it is not absolutely clear that field theory remains a useful approximation. Nor is there any sensible measure on the space of initial conditions which would justify the expectation that macroscopic regions could inflate with non-neglible probability.

A second troublesome feature is the supposition that $Q$ comes to dominante the universe only at recent redshifts. This is rather difficult to justify on general grounds. It is uncomfortable to be required to suppose that we are living at a privileged point in the universe's history when the details of this transition are visible. In epochs long ago vacuum energy would have been strongly subdominant, and an almost unobservable component of the matter mix. In epochs far in the future, vacuum energy will be highly dominant, and all other forms of matter will be comparatively invisible. Therefore there is a strong desire to find dynamical mechanisms, independent of the initial conditions, which make our existence at this privileged epoch understandable.

Thirdly, $Q$ must be very weakly coupled (Amendola and Tocchini-Valentini, 2001; Maccio, Quercellini, Mainini, Amendola, and Bonometto, 2003; Tocchini-Velentini and Amendola, 2002). This is necessary in order that quintessence could have evaded detection via particle physics accelerator experiments or other cosmological probes, such as nucleosynthesis, which measure the cross section for Standard Model particles to decay into exotic forms of matter. The constraints here are tight, so the probability that familiar particles such as baryons or leptons mutate into quintessence cannot be appreciable. Were we to suppose that quantum field theory were a fundamental description of nature, then it would be more or less acceptable to restrict our attention to renormalizable field theories in which the number of ways in which $Q$ could combine with Standard Model fields would be limited by the fundamental requirement that in $D$ dimensions the dimensionality of any operator in the Lagrangian should be $\leqslant D$. Unfortunately this point of view has long since been abandoned, and we now expect any effective four-dimensional field theory to descend, essentially via the Wilsonian renormalization group, from a more fundamental supergravity, string- or M-theory description of nature. In that case one must expect the effective Lagrangian to contain an infinite number of terms of arbitrary dimensionality $d > D$, each suppressed by a power $M^{D-d}$ of the mass scale $M$ at which the effective four-dimensional

theory fails. This means, for example, that one generically expects couplings of the form

$$\mathcal{L}_1 = \beta \frac{F^2 Q}{M} \tag{6.1.2}$$

where $F = \mathrm{d}A$ is the Maxwell field strength, $Q$ is the quintessence field, and $\beta$ is a dimensionless coupling which one should expect on naturalness grounds to be of order unity. This term will appear accompanied by an infinite number of other combinations which provide any number of ways for normal matter to transform into quintessence over the history of the universe. Although one must evaluate the bounds arising from couplings of this type carefully, in order to properly assess the magnitude of the problem, it is clear that if a quintessence field is the method nature has chosen to provide acceleration, then some other mechanism must be operating which suppresses the couplings $\beta$.

## 6.2. Quintessence in the braneworld

In the braneworld very much the same considerations apply that were outlined in the previous section. In particular, one expects the low energy theory describing the brane universe to arise from the dimensional reduction of a more fundamental string- or M-theory description, by integrating out physics above the energy scale at which the extra dimensions become visible. For the braneworld, the natural supposition is that the cut-off scale $M$ above which new physics appears coincides with the Planck scale of the higher-dimensional theory, possibly of order a TeV or so. Above $M$, the theory is no longer well approximated by a four-dimensional theory, and we need the details of extra transverse dimensions.

The coupling (6.1.2) describes interactions between the Maxwell field and quintessence, and will lead to cosmic variation of the fine structure constant (Carroll, 1998; Parkinson, Bassett, and Barrow, 2003). More generally, given a coupling $\mathcal{L}_n = \beta_n Q^n \mathcal{L}_{(4)}/M^n$, where $\mathcal{L}_{(4)}$ is any dimension 4 electromagnetic operator, and assuming that $Q$ varies only slowly, gives

$$\Delta\alpha \simeq -n\beta^n Q^{n-1}|_{\overline{Q}} \frac{\Delta Q}{M^n} \tag{6.2.1}$$

where $\overline{Q} = Q(0)$, the value of $Q$ today; the symbol $\Delta Q$ abbreviates the field interval $\Delta Q(z) = Q(z) - Q(0)$; and $z$ represents the redshift dependence of $Q$.

## 6.3. Radiative corrections to quintessence couplings and masses

The arguments outlined above give restrictions on a given set of couplings constants and shape parameters for some popular, rather generic, potentials. In doing so we have implicitly assumed that the form of the potential can be given as an Ansatz, so strictly speaking we were dealing with renormalized quantities, and the potential we were describing was the quantum effective potential (Weinberg, 1994). A more subtle question is to ask how a given tree-level potential is modified when quantum effects are taken into account. This involves the study of loop corrections to both the quintessence potential and the coupling of $Q$ to other fields.

For scalar quintessence coupled to a fermion species, this was first done by Doran and Jäckel (2003). Horvat (2002) has considered coupling to neutrinos, but the conclusions in this case do not strongly depend on physics modified by the braneworld scenario, so we shall not pay much attention to the neutrino example. Instead, we briefly review coupling to fermions, and explain how to translate the principal results to a brane scenario.

Since we are dealing with gravitational processes, it is mostly convenient to work in the Euclidean picture, which is obtained by complexifying the metric. Suppose a spacetime interval in $(D+1)$ dimensions is described by the displacement

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \hat{g}_{ij}\mathrm{d}x^i\mathrm{d}x^j \tag{6.3.1}$$

with $t$ a timelike coordinate, and $\hat{g}_{ij}$ a positive definite $D$ dimensional metric. One usually insists that $t$ is restricted to real values only, but instead suppose that $t$ is allowed to range over all of $\mathbf{C}$. The result is said to the complexification (in time) of the original metric. By restricting $t$ to specific contours in $\mathbf{C}$ one is said to obtain various sections of the metric: the Lorentzian section—where $t$ follows the real axis—recovers the original metric, whereas the Euclidean section is the case where $t$ is restricted to the imaginary axis. In the quintessence sector, we work with the Euclidean action

$$S = \int \mathrm{d}^4x\, \sqrt{-g}\left(\frac{1}{2}\partial_a Q\partial^a Q + V(Q) + \bar{\psi}[\partial\!\!\!/ + m(Q)]\psi\right) \tag{6.3.2}$$

where $Q$ is the quintessence field, $V(Q)$ is its classical potential, $\psi$ is a four-dimensional Dirac fermion, and $m(Q)$ a possibly field-dependent mass term for the fermion particle. In the braneworld, the integral is over the four-dimensional worldvolume of the brane which carries our universe.

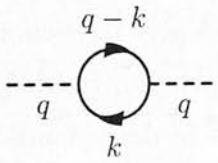



**Figure 6.1.** Lowest order fermion loop contributing to the effective quintessence potential. The dashed lines represent quintessence particles, and solid lines are fermons.

The classical potential only explicitly describes the interactions and couplings between particles which appear at tree level, but in the quantum theory extra couplings may be generated and existing couplings modified by radiative effects. These are processes in which an in-state containing some number of particles may be linked to an out-state through any number of intermediate states with arbitrary particle content. The net result is an effective coupling between the particles in the out-state and the particles in the in-state, even where no such coupling is present in the classical potential. The particles present in intermediate states are not physical in any meaningful sense, and are usually known as virtual particles. Most importantly, since all particles couple in the same way to gravity, there is always a minimal level of coupling between any two particle species which is mediated by the exchange of virtual gravitons. We shall explore such gravitational couplings later, in Section 6.6. The leading correction to the quintessence potential in the fermionic sector comes from the diagram of Fig. 6.1, in which two fermions circulate in a loop between an in-state and an out-state each consisting of a single quintessence particle.

**6.3.1. Constraints on the mass term $m(Q)$.** The fermion mass term $m(Q)$ is the source of some rather strong constraints on the possible behaviour of the quintessence field $Q$. If we are going to ask that the form of the quintessence potential is stable against radiative corrections, then to a first approximation it is sufficient to demand that the loop correction $V^*$ is small in comparison to the classical potential, $V^* \ll V$. This is the source of most of the bounds we calculate in this chapter.

The calculation of the quantum effective potential has already been discussed in Chapter 2. It is obtained by calculating the vacuum–vacuum amplitude with a shifted field,[1]

$$\Gamma[Q_0] = \int_{\substack{\text{1PI}\\ \text{connected}}} [\mathrm{d}Q]\ \exp(\mathrm{i}I[Q+Q_0]), \tag{6.3.3}$$

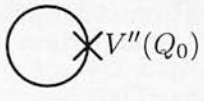
where the integral is restricted to one particle irreducible (1PI) connected diagrams. This is the effective action. If the fields are independent of position then every term in $\Gamma[Q_0]$ will contain a factor of the volume of spacetime, $V_4 = (2\pi)^4\delta(0)$. The quantum effective potential is the remainder of $\Gamma[Q_0]$ when this volume has been divided out. In the present case, one obtains

$$\Gamma[Q_0] = \int_{\substack{\text{1PI}\\ \text{connected}}} [\mathrm{d}Q][\mathrm{d}\psi] \exp\left(\mathrm{i}\int \mathrm{d}^4x\, \frac{1}{2}\partial^a Q \partial_a Q + V(Q_0) + \frac{1}{2}V''(Q_0)Q^2 + \bar{\psi}\left[\not{\partial} + m(Q_0)\right]\psi\right), \tag{6.3.4}$$

We have expanded the potential $V(Q)$ to second order, although in principle an infinite tower of couplings will be generated in this way. The first-order piece $V(Q)$ is missing because the classical configuration $Q_0$ is assumed to sit at a minimum of the tree level potential, for which $V'(Q_0) = 0$. If this is not true then we are expanding around an unstable vacuum and the resulting field theory will usually exhibit a number of pathologies. We have disregarded a coupling between $Q$ and $\psi$ and evaluated the result for Minkowski space, in which a non-trivial volume Jacobian or couplings arising via the spin connexion are absent. The 0-loop term is the $Q$ independent piece, which can be factored out immediately; it only gives back $V(Q_0)$. The 1-loop correction is given by the 1-loop piece in the vacuum–vacuum amplitude

$$\int_{\substack{\text{1PI}\\ \text{connected}}} [\mathrm{d}Q][\mathrm{d}\psi] \exp\left(\mathrm{i}\int \mathrm{d}^4x\, \frac{1}{2}\partial^a Q \partial_a Q + \frac{1}{2}V''(Q_0)Q^2 + \bar{\psi}\left[\not{\partial} + m(Q_0)\right]\psi\right). \tag{6.3.5}$$

The lowest order pure scalar loop has only the effective mass term $V''(Q_0)$,

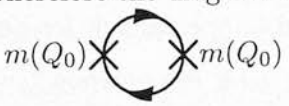



The amplitude for this diagram is

$$\int \mathrm{d}^4p \frac{-\mathrm{i}}{(2\pi)^4}\frac{1}{p^2-\mathrm{i}\varepsilon}(-\mathrm{i})(2\pi)^4 V''(Q_0)\delta_{\mathrm{D}}^{(4)}(p-p) = \frac{V_4}{(2\pi)^4}V''(Q_0)\pi^2\int_0^\infty p\,\mathrm{d}p \sim \frac{V_4}{32\pi^2}V''(Q_0)\Lambda^2 \tag{6.3.6}$$

[1]There is no factor of i in the definition of $\Gamma[Q_0]$ because we are already dealing with Euclidean quantities.

where $\Lambda$ is an explicit ultra-violet cut-off. One might expect the lowest-order fermion loop to be of the same sort. While this is roughly true, the lowest-order loop has two insertions corresponding to the mass self-coupling $m(Q_0)$ because the loop with only one insertion vanishes on symmetry grounds. Therefore the diagram has the form




The amplitude here is

$$-\operatorname{Tr}\int \mathrm{d}^4p\,\delta(p-p)\frac{-\mathrm{i}}{(2\pi)^4}\frac{-\mathrm{i}\not{p}^A{}_B}{p^2-\mathrm{i}\varepsilon}(2\pi)^4(-\mathrm{i})m(Q_0)\mathbf{1}^B{}_C\frac{-\mathrm{i}}{(2\pi)^4}\frac{-\mathrm{i}\not{p}^C{}_D}{p^2-\mathrm{i}\varepsilon}(2\pi)^4(-\mathrm{i})m(Q_0)\mathbf{1}^D{}_E \tag{6.3.7}$$

where $\mathbf{1}$ is the unit matrix in the Dirac algebra and the trace is over Dirac indices. The product $\not{p}^2$ evaluates to $p^2$, so this is just the same as the scalar loop except for the factor $\operatorname{Tr}\mathbf{1} = 4$ coming from the Dirac terms and the leading minus sign appropriate to a fermion loop. Therefore the 1-loop effective potential is

$$V^*(Q_0) = V(Q_0) + \frac{\Lambda^2}{32\pi^2}V''(Q_0) - \frac{\Lambda^2_{\mathrm{ferm}}}{8\pi^2}m(Q_0)^2, \tag{6.3.8}$$

where we have left open the possibility of an independent cut-off $\Lambda_{\mathrm{ferm}}$ for the fermion species. The restriction $V^*/V \ll 1$ is a condition on the numerical value of both the correction terms to $V(Q_0)$.

The scalar term just re-expresses the fact that the potential must be flat. In the fermion term, suppose for definiteness that $m(Q)$ is determined by some large field independent bare mass $m_0$ plus a correction term $c$ which is generated by couplings to other fields, in this case $Q$.

$$m(Q) = m_0 + c(Q) \tag{6.3.9}$$

The field-independent piece $m_0$ does not contribute to the couplings of $Q$, so requiring that the loop corrections is small by comparison with the classical piece gives a condition[2]

$$\frac{\Lambda^2_{\mathrm{ferm}}}{4\pi^2}\frac{m_0c(Q_0)}{V(Q_0)} \ll 1 \quad \text{or,} \quad c \ll \frac{4\pi^2}{\Lambda^2_{\mathrm{ferm}}}\frac{V(Q_0)}{m_0}, \tag{6.3.10}$$

where we have neglected the second order term in $c$ because of our expectation that the mass correction will turn out to be small. This just arises from demanding that the third

[2]This is not exact, because the 1-loop effective potential should now be calculated with a mass term corresponding to $m_0$ in the fermion propagator, but (6.3.8) still gives a useful estimate of the constraint. In following this technique, we are replicating the argument of Doran and Jäckel (2003).

term in (6.3.8) is small in comparison with the third term, and we have dropped the field-independent contribution from $m(Q_0)^2$ which is proportional to $m_0^2$, since this does not affect the *shape* of the potential. The bound (6.3.10) strongly constrains the field-mediated mass $c(Q_0)$ because the fermion cut-off $\Lambda$ is usually rather large, perhaps of order the four-dimensional Planck scale $M_{\rm P} = 10^{19}$ GeV, or the GUT scale at a few orders of magnitude less. This is the first example of how a modification in the Planck scale, from $10^{19}$ GeV to (say) a TeV could give rise to very large modifications in the kind of constraints we are dealing with.

In the present case to obtain a proper numerical estimate one must supply a value for the present-day potential $V(Q_0)$. Although there is no a priori way to ascertain the absolute value of $V(Q_0)$, except perhaps by computing it from first principles using a more fundamental theory (a proposition far out of reach with present day technology) one can make a gross estimate by supposing that $Q$ dominates the present energy density of the universe. Since our observable patch of the universe is apparently very nearly flat, one has

$$H^2 = \frac{\kappa_4^2}{3} V(Q_0). \tag{6.3.11}$$

This means that $V(Q_0)$, at least in our vicinity, must be comparable to the critical density

$$\rho_{\rm crit} \simeq 8.1 \times 10^{-11}\, h^2\ {\rm eV}^4. \tag{6.3.12}$$

If we set $\Lambda$ to be around the GUT scale ($10^{-3} M_{\rm P}$) and take the bare mass to be around the supersymmetry breaking scale (of order a TeV), the constraint we derive is rather severe:

$$c \ll 10^{-71}\ {\rm eV}. \tag{6.3.13}$$

Notice that this justifies our neglect of $c^2$ terms.

There is a coupling of the gravitational field, represented by weak perturbations of the metric $g_{ab}$, to any form of matter via the volume Jacobian $\det g_{ab}$ which unavoidably appears when constructing an action from a scalar Lagrangian density. Although gravity couples only very weakly to other forms of matter, the severity of (6.3.13) gives rise to some concern that gravitationally mediated couplings between $\psi$ and $Q$ could induce a fermion mass which violates this $c$-bound. The results of Doran and Jäckel (2003) demonstrate that this does not happen in four dimensions, but the braneworld case remains open.

The calculation leading to Eq. (6.3.13) depends only on the details of quantum field theory in the four-dimensional world, so it is valid on the worldvolume of a brane universe

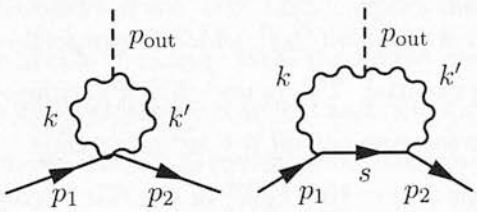



**Figure 6.2.** Low-order gravitationally mediated diagrams contributing to the quintessence–fermion coupling. Fermions are solid lines; gravitons are wavy lines. Quintessence is shown as dashes, and since the graviton couples to the entire quintessence potential, this in principle involves arbitrary powers of $Q$.

provided that we take the effective cut-off $\Lambda$ to be sized appropriately. Since the bound on $c$ scales with $\Lambda^{-2}$, it follows that a reduction in the scale of the cut-off will weaken any constraint: it will be easier for any given model to satisfy the bound on radiatively induced masses. For example, in a model where $\Lambda$ should be of order a TeV $\sim 10^{12}$ eV, one finds that $c \ll 10^{-44}$ eV. This weakening is a mixed blessing. It is harder to rule out any given model of quintessence, but it may make the construction of attractive phenomenological models easier.

We now turn to the concrete calculational task of estimating the magnitude of a $\bar{\psi}\psi Q$ type coupling in the effective Lagrangian. We assume that at tree level there is no bare coupling, in which case the leading order diagrams contributing to the radiatively induced coupling are shown in Fig. 6.2. An evaluation of these diagrams rests on the Feynman rules, the vertices and propagators of which can be read off or calculated from the Lagrangian in the usual way. In doing so one must remember that the four-dimensional matter theories are confined to the worldvolume of the brane, and therefore enter with a $\delta$-function which restricts their support. The gravitational theory is properly five-dimensional and behaves as one would expect. Because of the presence of $\delta$-functions, it is convenient to formulate the Feynman rules in coordinate space rather than momentum space. This is not essential, but we shall see that the ease of calculation greatly recommends this approach.

## 6.4. The gravitational propagator

We work with Minkowski branes. Although this approximation is rather crude, it is useful for a first attempt at the problem and facilitates comparison with earlier work

which also makes use of the Minkowski idealization. The vertices and the quintessence and fermion propagators can be simply read off the Lagrangian, but it is more work to calculate the graviton propagator. In this section we carry out the calculation, and use the result in the next.

The graviton propagator was first derived by Giddings et al. (2000), who solved for the gravitational Green's function. Although this is sufficient, one should properly fix the gauge using the Fadeev–Popov procedure. This is the approach that we shall follow, although the final answers will agree. We adopt the conventional brane line element (Binetruy et al., 2000a,b), as described in (5.1.1),

$$\mathrm{d}s^2 = g_{ab}\mathrm{d}x^a\mathrm{d}x^b = -n^2(t,y)\mathrm{d}t^2 + a^2(t,y)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j + \mathrm{d}y^2 \tag{6.4.1}$$

where $y$ is a Gaussian normal coordinate transverse to the brane. This metric is taken to be a solution of the five-dimensional Einstein equations with cosmological constant $\Lambda$, but vanishing bulk energy–momentum tensor. The brane is embedded at $y = 0$ and there is a $\mathbf{Z}_2$ symmetry which acts on the Gaussian normal coordinate as $y \mapsto -y$. As was described in Section 5.1.1, there is a coordinate horizon where the Gaussian normal coordinate used to write the metric in this form breaks down, and we write the location of this horizon as $y = y_\mathrm{h}$.

Gravitational disturbances were discussed in Section 5.7 and take the form of small perturbations to the metric, written $h_{ab}$

$$\mathrm{d}s^2 = (g_{ab} + h_{ab})\mathrm{d}x^a\mathrm{d}x^b, \tag{6.4.2}$$

where $h_{ab}$ is first order in the sense that each component is taken to satisfy $|h_{ab}| \ll 1$. In a general $D$-dimensional spacetime without isometries, it is simple in explicit calculations to treat each component of $h_{ab}$ as a separate scalar field. However this approach is inefficient and leads to a needless profusion of separate scalars where a number of isometries exist in the background. In the present case, one can decompose $h_{ab}$ into representations of the brane isometry group which consists of rotations and translations along the spacelike directions of the brane's worldvolume. In Minkowski space, one can enlarge this symmetry group to include transformations which mix the timelike direction.

At the outset, we fix part of the gauge freedom by demanding that $y$ remains a Gaussian normal coordinate. Therefore the $h_{ay}$ components of graviton, for any label $a$, must be zero. Since we are dealing with Minkowski branes, the remaining pieces of $h_{ab}$ form a

representation of the brane isometry group on their own and can profitably be left intact. The remaining gauge freedom in this piece is what will be fixed by the Fadeev–Popov procedure. More generally, the remainder would split into a spin-2 representation of the spacelike isometry group, two spin-1 representations and a spin-0 piece,

$$(\mathrm{d}x^i \otimes \mathrm{d}x^j) \oplus (\mathrm{d}t \otimes \mathrm{d}x^j) \oplus (\mathrm{d}x^i \otimes \mathrm{d}t) \oplus (\mathrm{d}t \otimes \mathrm{d}t) = \mathrm{d}x^a \otimes \mathrm{d}x^b \tag{6.4.3}$$

where here $a$, $b$ take values in $t$, $i$, ..., but not $y$. What we are going to calculate in this section is the propagator for the spin-2 piece. In the special case where the brane carries an Minkowski cosmology, which we shall specialise to later in the derivation, this spin-2 piece is all that is required to carry full information for the graviton.[3]

**6.4.1. The Einstein–Hilbert action.** To this end, we agree to deal with a perturbed metric of the form (cf. (5.7.1))

$$\mathrm{d}s^2 = -n^2(t,y)\,\mathrm{d}t^2 + a^2(t,y)(\delta_{ij} + e_{ij})\mathrm{d}x^i\mathrm{d}x^j + \mathrm{d}y^2 \tag{6.4.4}$$

where we have factored $a^2$ out of the graviton field are, as described above, are dealing only with the spin-2 piece under the spacelike brane isometry group. The purely bulk gravitational action is

$$S_g = -\frac{1}{2\kappa_5^2}\int \mathrm{d}^5x\,\sqrt{-\det g}(R + 2\Lambda), \tag{6.4.5}$$

to which one must add a brane matter theory supporting fermions and a quintessece field $Q$, but for the present these terms are not yet necessary. The components of the connexion are

$$\Gamma^t_{tt} = \frac{\dot{n}}{n}, \quad \Gamma^t_{ij} = \frac{a\dot{a}}{n^2}(\delta_{ij} + e_{ij}) + \frac{a^2}{2n^2}\dot{e}_{ij}, \quad \Gamma^t_{yt} = \frac{n'}{n} \tag{6.4.6a}$$

$$\Gamma^y_{tt} = nn', \quad \Gamma^y_{ij} = -aa'(\delta_{ij} + e_{ij}) - \frac{a^2}{2}e'_{ij} \tag{6.4.6b}$$

$$\Gamma^j_{tk} = \frac{\dot{a}}{a}\delta^j_k + \frac{1}{2}(\delta^{ij} - e^{ij})\dot{e}_{ik}, \quad \Gamma^j_{yk} = \frac{a'}{a}\delta^j_k + \frac{1}{2}(\delta^{ij} - e^{ij})e'_{ik} \tag{6.4.6c}$$

$$\Gamma^j_{km} = (\delta^{ij} - e^{ij})\partial_{(k}e_{m)i} - \frac{1}{2}(\delta^{ij} - e^{ij})\partial_i e_{km}, \tag{6.4.6d}$$

and the Ricci scalar is

$$R = R_0 + R_1 + R_2, \tag{6.4.7}$$

[3]One needs to include the other Kaluza–Klein fields of lower spins when the background does not have the full isometries of Minkowski space.

where the zero, first and second order pieces are

$$R_0 = -2\frac{n''}{n} + \frac{6}{n^2}\frac{\ddot{a}}{a} - 6\frac{a''}{a} - \frac{6}{n^2}\frac{\dot{a}}{a}\frac{\dot{n}}{n} - 6\frac{a'}{a}\frac{n'}{n} + 6\left(\frac{\dot{a}}{a}\right)^2 - 6\left(\frac{a'}{a}\right)^2 \tag{6.4.8a}$$

$$R_1 = \frac{1}{n^2}\ddot{e} - e'' + \frac{1}{n^2}\dot{e}\left(4\frac{\dot{a}}{a} - \frac{\dot{n}}{n}\right) - e'\left(4\frac{a'}{a} + \frac{n'}{n}\right) + \frac{1}{a^2}\partial_i\partial_j e^{ij} - \frac{1}{a^2}\triangle e \tag{6.4.8b}$$

$$\begin{aligned} R_2 = &-\frac{1}{n^2}e^{ij}\ddot{e}_{ij} + e^{ij}e''_{ij} - \frac{3}{4}\frac{1}{n^2}\dot{e}_{ij}\dot{e}^{ij} + \frac{3}{4}e'_{ij}e'^{ij} \\ &+ \frac{1}{n^2}e^{ij}\dot{e}_{ij}\left(\frac{\dot{n}}{n} - 4\frac{\dot{a}}{a}\right) + e^{ij}e'_{ij}\left(\frac{n'}{n} + 4\frac{a'}{a}\right) + \frac{1}{4}\frac{1}{n^2}\dot{e}\dot{e} - \frac{1}{4}e'e' \\ &- \frac{2}{a^2}e^{ij}\partial_k\partial_i e_j{}^k + \frac{1}{a^2}e^{ij}\triangle e_{ij} + \frac{1}{a^2}e^{ij}\partial_i\partial_j e - \frac{1}{a^2}\partial_i e^{ik}\partial_k e^k{}_j \\ &+ \frac{3}{4}\frac{1}{a^2}\partial_k e^{ij}\partial^k e_{ij} - \frac{1}{4}\frac{1}{a^2}\partial_i e\partial^i e - \frac{1}{2}\frac{1}{a^2}\partial_i e_{jk}\partial^k e^{ij} + \frac{1}{a^2}\partial_i e^{ij}\partial_j e \end{aligned} \tag{6.4.8c}$$

where $e = \mathrm{Tr}\, e = \delta^{ij}e_{ij}$ and we have used the fact that $e_{ij}$ and $e^{ij}$ are numerically equal, since indices are raised and lowered via the flat Euclidean metric. The remaining piece necessary to construct the Lagrangian is the invariant volume element $\sqrt{-g}$, which must be computed to second order. This is most easily evaluated using the matrix identity $\ln\det\mathbf{m} = \mathrm{Tr}\ln\mathbf{m}$ and elementary properties of determinants,

$$\begin{aligned} \det g &= \det\begin{pmatrix} -n^2 & & \\ & a^2(\delta_{ij}+e_{ij}) & \\ & & 1 \end{pmatrix} = \det\begin{pmatrix} n^2 & & \\ & a^2\delta_{ij} & \\ & & \end{pmatrix}\begin{pmatrix} 1 & & \\ & & \\ & \delta_{ij}+e_{ij} & \end{pmatrix} \\ &= -n^2a^6\exp\mathrm{Tr}\ln\begin{pmatrix} 1 & & \\ & & \\ & \delta_{ij}+e_{ij} & \end{pmatrix} \end{aligned} \tag{6.4.9}$$

The matrix logarithm is defined by its power series, $\ln(\mathbf{1}+\mathbf{m}) \simeq \mathbf{m} - \mathbf{m}^2/2$ (to second order), so

$$\sqrt{-\det g} = na^3\left(1 + \frac{e}{2} - \frac{e^{ij}e_{ij}}{4} + \frac{e^2}{8}\right) \tag{6.4.10}$$

Putting all these elements together, it is now possible to compute the action. Picking out the quadratic part gives[4]

$$S_g = -\frac{1}{2\kappa_5^2}\int \mathrm{d}^5x\; na^3\left(L_\Lambda + L_t + L_y + L_\partial\right), \tag{6.4.11}$$

[4]The terms proportional to a single power of $e$ must vanish, because we are expanding around an on-shell solution. Therefore they can justifiably be ignored, which results in a gratifying simplification of the algebra. In principle this would provide a test of the correctness of the reduction of the action, but in practice the difficulty of carrying along all first order terms generates more errors than it corrects.

where $L_\Lambda$ is a contribution proportional to the cosmological constant, $L_t$ involves $t$ derivatives, $L_y$ involves $y$ derivatives, and $L_\partial$ is spatial derivative piece,[5]

$$L_\Lambda = (R_0 + 2\Lambda)\left(\frac{e^2}{8} - \frac{e^{ij}e_{ij}}{4}\right) \tag{6.4.16a}$$

$$L_t = \frac{1}{2n^2}e\ddot{e} + \frac{1}{4n^2}\dot{e}\dot{e} - \frac{1}{2n^2}e\dot{e}\left(\frac{\dot{n}}{n} - 4\frac{\dot{a}}{a}\right) - \frac{1}{n^2}e^{ij}\ddot{e}_{ij} - \frac{3}{4n^2}\dot{e}_{ij}\dot{e}^{ij} + \frac{1}{n^2}e^{ij}\dot{e}_{ij}\left(\frac{\dot{n}}{n} - 4\frac{\dot{a}}{a}\right) \tag{6.4.16b}$$

$$L_y = -\frac{1}{2}ee'' - \frac{1}{4}e'e' - \frac{1}{2}ee'\left(\frac{n'}{n} + 4\frac{a'}{a}\right) + e^{ij}e'_{ij} + \frac{3}{4}e'_{ij}e'^{ij} + e^{ij}e'_{ij}\left(\frac{n'}{n} + 4\frac{a'}{a}\right) \tag{6.4.16c}$$

$$\begin{aligned} L_\partial = {} & \frac{3}{2a^2}e^{ij}\partial_i\partial_j e - \frac{1}{2a^2}e\triangle e - \frac{2}{a^2}e^{ij}\partial_k\partial_i e_j{}^k + \frac{1}{a^2}e_{ij}\triangle e^{ij} - \frac{1}{a^2}\partial_i e^{ij}\partial_k e^k{}_j \\ & + \frac{3}{4a^2}\partial_k e_{ij}\partial^k e^{ij} - \frac{1}{4a^2}\partial_i e\partial^i e - \frac{1}{2a^2}\partial_i e_{jk}\partial^k e^{ij} + \frac{1}{a^2}\partial_i e^{ij}\partial_j e \end{aligned} \tag{6.4.16d}$$

The background Ricci scalar $R_0$ is determined on-shell by the Einstein equations,

$$R_{ab} - \frac{1}{2}Rg_{ab} = \Lambda g_{ab}, \quad \text{so} \quad R_0 - \frac{5}{2}R_0 = 5\Lambda, \quad \text{or} \quad R_0 = -\frac{10\Lambda}{3}. \tag{6.4.17}$$

The $L_\Lambda$ sector therefore simplifies to give

$$L_\Lambda = -\frac{4\Lambda}{3}\left(\frac{e^2}{8} - \frac{e^{ij}e_{ij}}{4}\right). \tag{6.4.18}$$

This will be cancelled by other terms appearing in the action. To see this, it is first convenient to integrate by parts in order to obtain all terms in the canonical form $e\partial\partial e$, or

[5]The $L_\delta$ sector is the same as the action for a $n$-dimensional $SO(1, n-1)$ graviton,

$$\mathrm{d}s^2 = (\delta_{ab} + h_{ab})\mathrm{d}x^a\mathrm{d}x^b. \tag{6.4.12}$$

The Levi–Civita connexion following from this metric is

$$\Gamma^m_{ab} = \partial_{(a}h_{b)}{}^m - h^{mr}\partial_{(b}h_{a)r} - \frac{1}{2}\partial^m h_{ab} + \frac{1}{2}h^{mr}\partial_r h_{ab}, \tag{6.4.13}$$

and the action (after some calculation) comes out to be

$$S_g = -\frac{1}{2\kappa^2}\int \mathrm{d}^4x\,\left(\frac{1}{4}h^{ab}\triangle h_{ab} - \frac{1}{4}h\triangle h + \frac{1}{2}h^{ab}\partial_a\partial_b h - \frac{1}{2}h^{ab}\partial_c\partial_a h^c{}_b\right), \tag{6.4.14}$$

after expanding to second order and integrating by parts. This has the characteristic form $h^{ab}(Z_{abmn} - Z_{ambn})h^{mn}$, where $Z_{abmn}$ is the operator

$$Z_{abmn} = \delta_{am}\delta_{bn}\triangle + \frac{1}{2}\delta_{nb}\partial_a\partial_m. \tag{6.4.15}$$

This structure will reappear in the brane propagator.

(equivalently) in Gaussian form (cf. Section A.2). Principally we are concerned with the $\dot{e}_{ij}\dot{e}^{ij}$ and $e'_{ij}e'^{ij}$ terms,

$$\int \mathrm{d}^5x na^3 \left(-\frac{3}{4n^2}\dot{e}_{ij}\dot{e}^{ij}\right) = \int \mathrm{d}^5x \left[-\frac{3}{4}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{a^3}{n}\dot{e}_{ij}e^{ij}\right) + \frac{9}{4}\frac{a^2\dot{a}}{n}\dot{e}_{ij}e^{ij} + \frac{3}{4}\frac{a^3}{n}e^{ij}\ddot{e}_{ij} - \frac{3}{4}\frac{a^3}{n}\frac{\dot{n}}{n}e^{ij}\dot{e}_{ij}\right]. \tag{6.4.19a}$$

and

$$\int \mathrm{d}^5x\, na^3 \left(\frac{3}{4}e'_{ij}e'^{ij}\right) = \int \mathrm{d}^5x \left[\frac{3}{4}\frac{\mathrm{d}}{\mathrm{d}y}\left[a^3ne'_{ij}e^{ij}\right] - \frac{9}{4}a^2a'ne'_{ij}e^{ij} - \frac{3}{4}n'a^3e'_{ij}e^{ij} - \frac{3}{4}na^3e''_{ij}e^{ij}\right]. \tag{6.4.19b}$$

We discard surface terms arising from total $t$ or spatial derivatives, which can be justified by a standard argument. On the other hand, one should keep track of surface terms which arise at the brane, since these will not generally vanish. In this case there is a surface term,

$$\Sigma_g = -\frac{1}{2\kappa_5^2}\frac{3}{4}\int_{\partial M} \mathrm{d}^4x\, na^3e'_{ij}e^{ij}. \tag{6.4.20}$$

We can already anticipate that many of these surface terms will disappear in the special case of tensor modes, provided they contain $y$ derivatives at the brane. In view of the restrictive tensor boundary condition, Eq. (5.7.7). Using this procedure to rewrite $L_t$ and $L_y$ yields

$$L_t = -\frac{1}{4n^2}e^{ij}\ddot{e}_{ij} + \frac{1}{4n^2}e^{ij}\dot{e}_{ij}\left(\frac{\dot{n}}{n} - 7\frac{\dot{a}}{a}\right) + \mathrm{Tr}\, e \text{ terms} \tag{6.4.21a}$$

$$L_y = \frac{1}{4}e^{ij}e''_{ij} + \frac{1}{4}e^{ij}e'_{ij}\left(\frac{n'}{n} + 7\frac{a'}{a}\right) + \mathrm{Tr}\, e \text{ terms}, \tag{6.4.21b}$$

where we have suppressed $\mathrm{Tr}\, e$ terms to keep the working simple. One now separates part of the $e^{ij}e'_{ij}$ and $e^{ij}\dot{e}_{ij}$ terms and integrates by parts once more,

$$\int \mathrm{d}^5x\, na^3\left(-\frac{1}{n^2}\frac{\dot{a}}{a}e^{ij}\dot{e}_{ij}\right) = \int \mathrm{d}^5x \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(-\frac{a^3}{2n}\frac{\dot{a}}{a}e^{ij}e_{ij}\right) + \left(\frac{a^3}{n}\frac{\dot{a}}{a}\frac{\dot{a}}{a} + \frac{a^3}{2n}\frac{\ddot{a}}{a} - \frac{a^3}{2n}\frac{\dot{n}}{n}\frac{\dot{a}}{a}\right)e^{ij}e_{ij}\right] \tag{6.4.22a}$$

and

$$\int \mathrm{d}^5x\, na^3\left(\frac{a'}{a}e^{ij}e'_{ij}\right) = \int \mathrm{d}^5x \left[\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{a^3n}{2}\frac{a'}{a}e_{ij}e^{ij}\right) - \left(na^3\frac{a'}{a}\frac{a'}{a} + \frac{na^3}{2}\frac{n'}{n}\frac{a'}{a} + \frac{na^3}{2}\frac{a''}{a}\right)e^{ij}e_{ij}\right]. \tag{6.4.22b}$$

Collecting terms, we see that there is another surface term from the $y$ integrations,

$$\Sigma_g = -\frac{1}{2\kappa_5^2}\frac{1}{2}\int_{\partial M} \mathrm{d}^4x\, na^2a'e_{ij}e^{ij}, \tag{6.4.23}$$

and a total contribution proportional to $e_{ij}e^{ij}$ in the bulk which amounts to

$$\int \mathrm{d}^5x\ \frac{na^3}{2}e_{ij}e^{ij}\left[\frac{2}{n^2}\frac{\dot{a}}{a}\frac{\dot{a}}{a}+\frac{1}{n^2}\frac{\ddot{a}}{a}-\frac{1}{n^2}\frac{\dot{n}}{n}\frac{\dot{a}}{a}-2\frac{a'}{a}\frac{a'}{a}-\frac{n'}{n}\frac{a'}{a}-\frac{a''}{a}\right]. \tag{6.4.24}$$

To simplify this, it is possible to use the background field equations (essentially (5.1.3a)–(5.1.3c)),

$$\frac{1}{n^2}\frac{\ddot{a}}{a}-\frac{1}{n^2}\frac{\dot{n}}{n}\frac{\dot{a}}{a}-\frac{n'}{n}\frac{a'}{a}+\frac{1}{n^2}\frac{\dot{a}}{a}\frac{\dot{a}}{a}-\frac{a'}{a}\frac{a'}{a}=-\frac{\Lambda}{3} \tag{6.4.25a}$$

$$-\frac{a''}{a}+\frac{1}{n^2}\frac{\dot{a}}{a}\frac{\dot{a}}{a}-\frac{a'}{a}\frac{a'}{a}=-\frac{\Lambda}{3}+\frac{\lambda}{3}\delta_{\mathrm{D}}(y). \tag{6.4.25b}$$

Pairing these equations shows that the entire contribution from $e_{ij}e^{ij}$ terms becomes

$$-\int \mathrm{d}^5x\ na^3\frac{\Lambda}{3}e_{ij}e^{ij}+\int_{\partial M}\mathrm{d}^4x\ na^3\frac{\lambda}{3}e_{ij}e^{ij} \tag{6.4.25c}$$

The bulk part of this cancels the cosmological contribution proportional to $e_{ij}e^{ij}$ from $L_\Lambda$. One can carry out an exactly analogous procedure to reduce the $\mathrm{Tr}\,e$ terms which were omitted in the above calculation to canonical form. The result is the same, with the bulk piece cancelling the $e^2$ term in $L_\Lambda$. As a result, $L_\Lambda$ drops out of the action entirely. The result is

$$S_g=-\frac{1}{8\kappa_5^2}\int \mathrm{d}^5x\ na^3\ (L_t+L_y+L_\partial)+\Sigma_g, \tag{6.4.26}$$

where

$$L_t=-\frac{1}{n^2}e^{ij}\ddot{e}_{ij}+\frac{1}{n^2}e^{ij}\dot{e}_{ij}\left(\frac{\dot{n}}{n}-3\frac{\dot{a}}{a}\right)+\frac{1}{n^2}e\ddot{e}-\frac{1}{n^2}e\dot{e}\left(\frac{\dot{n}}{n}-3\frac{\dot{a}}{a}\right) \tag{6.4.27a}$$

$$L_y=e^{ij}e''_{ij}+e^{ij}e'_{ij}\left(\frac{n'}{n}+3\frac{a'}{a}\right)-ee''-ee'\left(\frac{n'}{n}+3\frac{a'}{a}\right) \tag{6.4.27b}$$

$$L_\partial=\frac{1}{a^2}e^{ij}\triangle e_{ij}-\frac{2}{a^2}e^{ij}\partial_k\partial_i e_j{}^k-\frac{1}{a^2}e\triangle e+\frac{2}{a^2}e^{ij}\partial_i\partial_j e, \tag{6.4.27c}$$

and $\Sigma_g$ is the cumulative surface term,

$$\Sigma_g=-\frac{1}{8\kappa_5^2}\int_{\partial M}\mathrm{d}^4x\ na^3\ \left(3e'_{ij}e^{ij}+2\frac{a'}{a}e_{ij}e^{ij}+ee'+\frac{a'}{a}e^2-\frac{4\lambda}{3}e_{ij}e^{ij}+\frac{2\lambda}{3}e^2\right). \tag{6.4.27d}$$

We will ignore this term in what follows. Although it is important, the self-interactions it describes do not contribute, in a first approximation, to the processes we intend to consider. At some point, one should come back to the surface term and properly incorporate its effects into the Feynman rules.

**6.4.2. Quantization of the graviton theory.** The quantization of theories like the theory of the $e_{ij}$ field was discussed at some length in Chapter 2. Since the action contains a gauge invariance, it must be properly gauge-fixed using the Fadeev–Popov procedure described in Section 2.4.2.

The structure of the action is somewhat verbose and it is very convenient to introduce an operator $\Box$ which describes the $t$ and $y$ derivatives,

$$\Box = \frac{1}{n^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial y^2} + \frac{\omega}{n^2}\frac{\partial}{\partial t} - \sigma\frac{\partial}{\partial y} \tag{6.4.28}$$

where

$$\omega = 3\frac{\dot{a}}{a} - \frac{\dot{n}}{n}, \quad \text{and} \quad \sigma = 3\frac{a'}{a} + \frac{n'}{n}; \tag{6.4.29}$$

and $\triangle$ is the $\delta_{ij}$ Laplacian. In terms of $\Box$, the action can be written

$$S_g = -\frac{1}{8\kappa_5^2}\int \mathrm{d}^5x\, na^3\, e^{ij}\left[(\delta_{ij}\delta_{rs} - \delta_{ir}\delta_{js})\Box + \frac{1}{a^2}(\delta_{ir}\delta_{js} - \delta_{ij}\delta_{rs})\triangle - \frac{2}{a^2}(\delta_{ir}\partial_j\partial_s - \delta_{ij}\partial_r\partial_s)\right] e_{ij}. \tag{6.4.30}$$

By factoring out the common index structure, this can be written somewhat more compactly as

$$S_g = -\frac{1}{8\kappa_5^2}\int \mathrm{d}^5x\, na^3\, e^{ij}\left[2\delta_{i[j}\delta_{r]s}\left(\Box - \frac{\triangle}{a^2}\right) + \frac{4}{a^2}\delta_{i[j}\partial_{r]}\partial_s\right] e_{rs}, \tag{6.4.31}$$

where brackets $[\cdots]$ denote antisymmetrization of total weight unity, ie., $2A_{[ij]} = A_{ij} - A_{ji}$. The spacelike brane isometry group now appears as an invariance of this action; Eq. (6.4.31) is invariant under transformations and rotations which mix the spacelike directions of the brane's worldvolume amongst themselves. This is a trivial consequence of writing the action in covariant form.

One passes to the quantum theory, as described in Chapter 2, by defining correlation functions of the field $e_{ij}$ using the functional integral. In the present case, this means that properly normalized, finite correlation functions of $e_{ij}$ are calculated using the rule

$$\langle e_{ij}(x)\cdots e_{mn}(y)\rangle = \frac{1}{\mathrm{Vol}(SO(3))}\int [\mathrm{d}e_{rs}]\; e_{ij}(x)\cdots e_{mn}(y)\; \exp \mathrm{i}S_g[e_{rs}]. \tag{6.4.32}$$

Since the action is invariant under $SO(3)$ transformations of $e_{ij}$, this Lagrangian is singular and we have divided by the volume of the gauge group in order to render the path integral finite. In Section 2.4.2 it was shown that this is equivalent to including an extra gauge-fixing piece in the action,

$$S \mapsto S + S_{\mathrm{gf}} = S - \frac{1}{2\kappa_5^2}\int \mathrm{d}^5x\, na^3\, \frac{1}{2\xi}(\partial e_b - \alpha\partial_b e)^2 \tag{6.4.33}$$

where $\xi$ and $\alpha$ are arbitrary numbers, and $\partial e_b - \alpha\partial_b e = \partial^a e_{ab} - \alpha\partial_b e^a{}_a$, with irrelevant indices suppressed to minimise clutter. In the language of Section 2.4.2, this says that the gauge fixing functional is $f_b = \partial^a e_{ab} - \alpha\partial_b e^a{}_a$, which is the analogue for general relativity of the Lorentz gauge condition in electromagnetism. In the limit where $\xi \to 0$, this enforces the transverse–traceless condition which is often called de Donder or Einstein gauge.

In writing these formulas, we are considering $e_{ab}$ to be a full five-dimensional tensor which is zero on $t$ and $y$ indices; where derivatives are contracted with $e_{ab}$ this makes no difference, but where two derivatives become contracted with themselves one must include contributions from the $t$ and $y$ sectors. Expanding this gauge-fixing piece out shows that $S_{\rm gf}$ satisfies

$$S_{\rm gf} = -\frac{1}{2\kappa_5^2}\frac{1}{2\xi}\int \mathrm{d}^5x\; na^3\; \left(\partial^a e_{ab}\partial^c e_c{}^b - 2\alpha\partial^a e_{ab}\partial_b e + \alpha^2\partial_b e\partial^b e\right). \tag{6.4.34}$$

After integrating by parts to obtain the action in the canonical form, that gives

$$S_{\rm gf} = -\frac{1}{2\kappa_5^2}\frac{1}{2\xi}\int \mathrm{d}^5x\; na^3\; e^{ij}\left(\frac{1}{a^2}\delta_{ir}\partial_j\partial_s - \frac{2\alpha}{a^2}\delta_{ij}\partial_r\partial_s + \frac{\alpha^2}{a^2}\delta_{ij}\delta_{rs}\triangle - \alpha^2\delta_{ij}\delta_{rs}\Box\right)e^{rs}. \tag{6.4.35}$$

Collecting all pieces together, we can write

$$S = S_g + S_{\rm gf} = -\frac{1}{2\kappa_5^2}\int \mathrm{d}^5x\; na^3\; e^{ij}\mathsf{D}_{ijrs}e_{rs}, \tag{6.4.36}$$

where the Gaussian kernel $\mathsf{D}$ satisfies

$$\mathsf{D}_{ijrs} = \left(\Box - \frac{\triangle}{a^2}\right)\left[-\frac{1}{4}\delta_{ir}\delta_{js} + \delta_{ij}\delta_{rs}\left(\frac{1}{4} - \frac{\alpha^2}{2\xi}\right)\right] - \frac{1}{a^2}\delta_{ir}\partial_j\partial_s\left(\frac{1}{2} - \frac{1}{2\xi}\right) + \frac{1}{a^2}\delta_{ij}\partial_r\partial_s\left(\frac{1}{2} - \frac{\alpha}{\xi}\right). \tag{6.4.37}$$

To simplify this result it is convenient to set $\xi = 1$ and $\alpha = 1/2$. With these choices the last two terms drop out, leaving the reduced operator

$$\mathsf{D}_{ijrs} = \left(\frac{1}{4}\delta_{i(r}\delta_{s)j} - \frac{1}{8}\delta_{ij}\delta_{rs}\right)\left(\Box - \frac{\triangle}{a^2}\right). \tag{6.4.38}$$

The propagator for the graviton field is the inverse of this operator,

$$\langle e_{ij}(x_1)e_{rs}(x_2)\rangle = -\mathrm{i}\kappa_5^2 G_{ijrs}(x_1, x_2) \tag{6.4.39}$$

where $\mathsf{D}_{ijrs}G^{rsmn}(x_1, x_2) = \delta^{(5)}_{\rm D}(x_1 - x_1)\delta^{(i}_r\delta^{j)}_s$.

In a quite general flat, homogeneous brane world the metric functions $n$ and $a$ are not equal, and both depend on $t$ and $y$. Therefore one has three spacelike Killing vectors $\partial/\partial x^i$ which generate translations along the spacelike coordinate axes, but no others. One

can exploit this symmetry by diagonalizing $\mathsf{D}_{ijrs}$ as a Fourier transform in the $x^i$, but, in general, the $t$ and $y$ dependence cannot be dealt with in the same manner. Proceeding in this fashion, one obtains (for a $d$-dimensional graviton)

$$G^{rsmn} = \left(\delta^{r(m}\delta^{n)s} - \frac{1}{d-2}\delta^{rs}\delta^{mn}\right)\int \frac{\mathrm{d}^3k}{(2\pi)^3}\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot(\mathbf{x}_1-\mathbf{x}_2)}G(t,y;\mathbf{k}), \tag{6.4.40}$$

where the momentum space function $G(t,y;\mathbf{k})$ obeys the equation

$$\frac{1}{4}\left(\Box + \frac{\mathbf{k}^2}{a^2}\right)G(t,y;\mathbf{k}) = \delta_{\mathrm{D}}(t_1-t_2)\delta_{\mathrm{D}}(y_1-y_2). \tag{6.4.41}$$

For the $SO(3)$ braneworld graviton, which is part of a multiplet with the vector $A_a$ and a scalar $\phi$, the dimension $d$ appearing in the group structure is 3. However in the Randall–Sundrum case, discussed below, we can enlarge $e_{ij}$ to an $SO(3,1)$ graviton, in which case $d = 4$.

**6.4.3. The Randall–Sundrum propagator.** At this point, one can make no further progress without specifying some explicit form for $n$ and $a$. The simplest choice is the Randall–Sundrum model described in Section 5.1.2, when the brane is taken to be empty of all matter, except for some intrinsic tension (Randall and Sundrum, 1999b) which is tuned to give a Minkowski brane. The line element is

$$\mathrm{d}s^2 = -\mathrm{e}^{-2\ell|y|}(-\mathrm{d}t^2 + \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j) + \mathrm{d}y^2. \tag{6.4.42}$$

The isometry group of the brane is $SO(3,1)$, the isometries of Minkowski space. In this case the derivation given above can be promoted to an $SO(3,1)$ graviton, rather the $SO(3)$ graviton plus vector and scalar pieces which were assumed above. This facilitates the comparison with four dimensions. The operator $\Box + \mathbf{k}^2/a^2$ is

$$\Box + \frac{k^2}{a^2} = \mathrm{e}^{2y\ell^{-1}}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial y^2} + \frac{4}{\ell}\frac{\partial}{\partial y} + \mathbf{k}^2\mathrm{e}^{2y\ell^{-1}} \quad (y>0), \tag{6.4.43}$$

and we assume $y > 0$ from now on. In this special case the functions $a$ and $n$ turn out to be equal, and all quantities are independent of the cosmic time $t$, so one recovers $\partial/\partial t$

as a Killing symmetry.[6] In many cases the presence of a symmetry corresponding to time translation, such as $\partial/\partial t$ is a great convenience in the description of physical phenomena, and this case is no exception. In particular, we are now entitled to diagonalize $G$ in the time direction, viz. $G(t, y; \mathbf{k}) = (2\pi)^{-1} \int \mathrm{d}\omega\, G(y; \omega, \mathbf{k}) \mathrm{e}^{\mathrm{i}\omega(t_1 - t_2)}$. This leaves only an ordinary differential equation for the $y$ dependence, which we might hope to solve by conventional means,

$$\left( -\frac{\partial^2}{\partial {y_1}^2} + \frac{4}{\ell} \frac{\partial}{\partial y_1} + (\mathbf{k}^2 - \omega^2) \mathrm{e}^{2y_1/\ell} \right) G(y; \omega, \mathbf{k}) = 4\delta_{\mathrm{D}}(y_1 - y_2). \tag{6.4.44}$$

Indeed, changing to a conformal bulk coordinate $z$ defined by $\mathrm{d}z = a\,\mathrm{d}y$ and setting $\beta^2 = \omega^2 - \mathbf{k}^2$, one obtains

$$\frac{z^2}{\ell^2} \left( \frac{\partial^2}{\partial {z_1}^2} - \frac{3}{z_1} \frac{\partial}{\partial z_1} + \beta^2 \right) G(z; \omega, \mathbf{k}) = -4 \frac{z_2}{\ell} \delta_{\mathrm{D}}(z_1 - z_2), \tag{6.4.45}$$

where we have used the well-known identity for the $\delta$-function of a function,

$$\delta_{\mathrm{D}}[g(x)] = \sum_i \frac{\delta_{\mathrm{D}}(x - x_i)}{|g'(x_i)|}, \tag{6.4.46}$$

in which the $x_i$ are the zeroes of $g(x)$. The quantity $\beta$ can be understood as (minus) the magnitude of a four-vector $k = (\omega, \mathbf{k})$, for which $k^2 = -\omega^2 + \mathbf{k}^2 = -\beta^2$ with our choice of signature. One now sets $G = z_1^2 \hat{G}$, which reduces the propagator equation to the Bessel equation,

$$\left[ \frac{\partial^2}{\partial {z_1}^2} + \frac{1}{z_1} \frac{\partial}{\partial z_1} + \left( \beta^2 - \frac{4}{z_1^2} \right) \right] \hat{G}(z; \omega, \mathbf{k}) = -4 \frac{\ell z_2}{z_1^4} \delta_{\mathrm{D}}(z_1 - z_2). \tag{6.4.47}$$

In the $z$-coordinate, the location of the brane is $z_{\mathrm{b}} = \ell$, but we can take its location to be arbitrary. When making numerical estimates, we restore $z_{\mathrm{b}} = \ell$.

From this point, the derivation coincides with the earlier derivation of Giddings et al. (2000). The general solution for $\hat{G}$ is any linearly independent combination of Bessel functions,

$$\hat{G} = AJ_2(\beta z_1) + BY_2(\beta z_1), \tag{6.4.48}$$

[6] Of course, this is the original Randall–Sundrum model. As was described in Chapter 5, the Randall–Sundrum model (and also the Kaloper–Linde branes, and a number of other brane metrics) can be given a bundle structure as a fibring of spacetime over a circle. Many of the special properties of the Randall–Sundrum model follow from this description, but the appearance of extra isometries such as $\partial/\partial t$ is not one of these. The isometries of the four-dimensional fibre metric can be chosen arbitrarily, and it is the fact that the Randall–Sundrum model represents Minkowski branes which is significant here.

where $A$ and $B$ are constants to be determined. Since this propagator represents tensor field $e_{ij}$, it should obey the boundary condition (5.7.7) (in the absence of anisotropic stress), so $\partial_y G|_{y=0} = 0$. The boundary condition in the far field can be obtained by analogy with electromagnetism, so we demand that postive frequency waves are purely ingoing as $y \to \infty$. (This is a causality condition that requires there is no information streaming in from infinity, which will destroy our solutions in the neighbourhood of the brane.) There are also the jump and continuity conditions characteristic of Green's functions,

$$[\hat{G}]^+_- = 0, \quad \text{and} \quad [(\partial/\partial z_1)\hat{G}]^+_- = -\frac{4\ell}{z_2^3}, \tag{6.4.49}$$

where the jump is taken across $z_1 = z_2$, and the last condition follows by integrating the propagator equation over a small neighbourhood of this surface. The boundary condition at the brane, in terms of $\hat{G}$, is

$$\frac{\partial}{\partial z_1}(z_1^2\hat{G})_{y=0} = 0, \quad \text{and so} \quad \left(2z_1\hat{G} + z_1^2\frac{\partial\hat{G}}{\partial z_1}\right)_{z_1=R} = 0, \tag{6.4.50}$$

where we have notionally placed the brane at $z = R$. That gives

$$A(2J_2 + \beta R J_2') + B(2Y_2 + \beta R Y_2') = 0, \tag{6.4.51}$$

where the arguments of the Bessel functions have been left unwritten and are all taken at $z = \beta R$. Using the Bessel recurrence relations

$$J_{\nu-1}(z) + J_{\nu+1}(z) = \frac{2\nu}{z}J_\nu(z) \tag{6.4.52a}$$

$$J_{\nu-1}(z) - J_{\nu+1}(z) = 2\frac{\mathrm{d}}{\mathrm{d}z}J_\nu(z), \tag{6.4.52b}$$

(where $J$ is any Bessel function), to rewrite $J_2'$ in terms of $J_1$ and $J_3$, and similarly for $Y_2'$. The remaining $J_2$, $Y_2$ terms can be replaced with combinations of $J_1$, $J_3$ and $Y_1$, $Y_3$, leaving

$$\beta R A J_1 + \beta R B Y_1 = 0, \tag{6.4.53}$$

in which all $J_3$, $Y_3$ dependence has cancelled out. Thus, by a redefinition of $A$, one has

$$G^- = A\left[Y_1(\beta R)J_2(\beta z_1) - J_1(\beta R)Y_2(\beta z_1)\right], \tag{6.4.54}$$

where we have written $G$ as $G^-$ in the region close to the brane, and $G^+$ in the far-field:

$$G(z_1, z_2; \omega, \mathbf{k}) = \begin{cases} G^-(z_1, z_2; \omega, \mathbf{k}) & R < z_1 < z_2 \\ G^+(z_1, z_2; \omega, \mathbf{k}) & z_1 > z_2 \end{cases}. \tag{6.4.55}$$

Implementing the far-field boundary condition is most easily done in terms of Hankel functions rather than Bessel and Neumann functions. The connexion is

$$H_\nu^{(1)}(z) = J_\nu(z) + \mathrm{i}Y_\nu(z), \quad \text{and} \quad H_\nu^{(2)}(z) = J_\nu(z) - \mathrm{i}Y_\nu(z). \tag{6.4.56}$$

Close to the brane, one has

$$G^- = \mathrm{i}A\left[J_1(\beta R)H_2^{(1)}(\beta z_1) - H_1^{(1)}(\beta R)J_2(\beta z_1)\right]. \tag{6.4.57}$$

The condition that positive frequency waves are purely ingoing at infinity implies that $G^+$ reduces to $H^{(1)}$,

$$G^+ = CH_2^{(1)}(\beta z_1) \tag{6.4.58}$$

for some $C$. The matching conditions at the boundary surface $z_1 = z_2$ are

$$\mathrm{i}A\left[J_1(\beta R)H_2^{(1)}(\beta z_2) - H_1^{(1)}(\beta R)J_2(\beta z_2)\right] = CH_2^{(1)}(\beta z_2) \tag{6.4.59a}$$

$$C\beta\frac{\mathrm{d}}{\mathrm{d}z_2}H_2^{(1)}(\beta z_2) - \mathrm{i}A\beta\left[J_1(\beta R)\frac{\mathrm{d}}{\mathrm{d}z_2}H_2^{(1)}(\beta z_2) - H_1^{(1)}(\beta R)\frac{\mathrm{d}}{\mathrm{d}z_2}J_2(\beta z_2)\right] = -\frac{4\ell}{z_2^3}. \tag{6.4.59b}$$

One replaces $C$ in the second equation with an expression in terms of $A$ found using the first. Two of the terms cancel, leaving

$$\mathrm{i}A\frac{H_1^{(1)}(\beta R)}{H_2^{(1)}(\beta z_2)}\left[H_2^{(1)}(\beta z_2)\frac{\mathrm{d}}{\mathrm{d}z_2}J_2(\beta z_2) - J_2(\beta z_2)\frac{\mathrm{d}}{\mathrm{d}z_2}H_2^{(1)}(\beta z_2)\right] = -\frac{4\ell}{\beta z_2^3}. \tag{6.4.60}$$

The term in square brackets is the Wronskian $\triangle[H_2^{(1)}, J_2]$ of $H_2^{(1)}$ and $J_2$. It is quite generally true that the Wronskian of any two linearly independent Bessel functions (say $J(z)$ and $Y(z)$) takes the general form $\triangle[J, Y] \propto z^{-1}$, with the constant of proportionality fixed by the choice of $J$ and $Y$. In this case,

$$\triangle[H_2^{(1)}(\beta z_2), J_2(\beta z_2)] = -\frac{2\mathrm{i}}{\pi\beta z_2}. \tag{6.4.61}$$

It's now easy to write $A$ and $C$ explicitly, using this expression for the Wronskian,

$$A = -\frac{2\pi\ell}{z_2^2}\frac{H_2^{(1)}(\beta z_2)}{H_1^{(1)}(\beta R)} \tag{6.4.62a}$$

$$C = -\frac{2\pi\mathrm{i}\ell}{z_2^2}\frac{1}{H_1^{(1)}(\beta R)}\left[J_1(\beta R)H_2^{(1)}(\beta z_2) - H_1^{(1)}(\beta R)J_2(\beta z_2)\right]. \tag{6.4.62b}$$

At this point it is possible to collect terms and write the full gravitational propagator,

$$G^{rsmn} = -2\pi\mathrm{i}\ell\left(\delta^{r(m}\delta^{n)s} - \frac{1}{2}\delta^{rs}\delta^{mn}\right)\int\frac{\mathrm{d}^3k\,\mathrm{d}\omega}{(2\pi)^4}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot(\mathbf{x}_1-\mathbf{x}_2)+\mathrm{i}\omega(t_1-t_2)}\frac{z_1^2}{z_2^2}\mathcal{G}, \tag{6.4.63}$$

where $\mathcal{G}$ satisfies

$$\mathcal{G} = \begin{cases} \dfrac{H_2^{(1)}(\beta z_2)}{H_1^{(1)}(\beta R)}\left[J_1(\beta R)H_2^{(1)}(\beta z_1) - H_1^{(1)}(\beta R)J_2(\beta z_1)\right] & R < z_1 < z_2 \\ \dfrac{H_2^{(1)}(\beta z_1)}{H_1^{(1)}(\beta R)}\left[J_1(\beta R)H_2^{(1)}(\beta z_2) - H_1^{(1)}(\beta R)J_2(\beta z_2)\right] & z_1 > z_2 \end{cases}. \tag{6.4.64}$$

Since the eventual interest is to evaluate the amplitude for processes happening on the brane, we shall be especially concerned with the special case in which both endpoints of the propagator lie on the brane, so that $z_1 = z_2 = R$. Using the identity

$$J_1(\beta R)H_2^{(1)}(\beta R) - H_1^{(1)}(\beta R)J_2(\beta R) = -\frac{2\mathrm{i}}{\pi\beta R} \tag{6.4.65}$$

permits the propagator to be rewritten,[7]

$$G^{rsmn}_{\text{brane}} = -\left(\delta^{r(m}\delta^{n)s} - \frac{1}{2}\delta^{rs}\delta^{mn}\right)\int \frac{\mathrm{d}^3k\,\mathrm{d}\omega}{(2\pi)^4}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot(\mathbf{x}_1-\mathbf{x}_2)+\mathrm{i}\omega(t_1-t_2)}\frac{4\ell}{\beta R}\frac{H_2^{(1)}(\beta R)}{H_1^{(1)}(\beta R)}. \tag{6.4.66}$$

This can be written in a manifestly Lorentz invariant form.

$$\begin{aligned} G^{rsmn}_{\text{brane}} &= -\left(\delta^{r(m}\delta^{n)s} - \frac{1}{2}\delta^{rs}\delta^{mn}\right)\int \frac{\mathrm{d}^4k}{(2\pi)^4}\mathrm{e}^{-\mathrm{i}k\cdot(x_1-x_2)}\frac{4\ell}{\mathrm{i}kR}\frac{H_2^{(1)}(\mathrm{i}kR)}{H_1^{(1)}(\mathrm{i}kR)} \\ &= -\left(\delta^{r(m}\delta^{n)s} - \frac{1}{2}\delta^{rs}\delta^{mn}\right)\int \frac{\mathrm{d}^4k}{(2\pi)^4}\mathrm{e}^{-\mathrm{i}k\cdot(x_1-x_2)}\frac{4\ell}{kR}\frac{K_2(kR)}{K_1(kR)}, \end{aligned} \tag{6.4.67}$$

where $K_\nu(z)$ is Macdonald or Basset function

$$K_\nu(z) = \frac{1}{2}\pi\mathrm{i}^{\nu+1}H_\nu^{(1)}(\mathrm{i}z). \tag{6.4.68}$$

## 6.5. The brane matter theory

Having successfully constructed the propagator in the gravitational sector, the next step is to couple gravity to a matter theory on the brane. The interesting theory here was presented in (6.3.2), and consists of a Dirac fermion $\psi$ and a scalar quintessence $Q$,

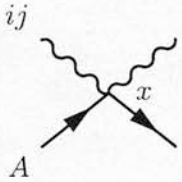
$$S_{\text{brane}} = \int \mathrm{d}^4x\,\sqrt{-\det h}\left[\frac{1}{2}\nabla_a Q\nabla^a Q + V(Q) + \bar{\psi}(\not{\nabla} + m)\psi\right]. \tag{6.5.1}$$

[7] If desired, this can be written in a way which makes the emergence of a zero mode and the Kaluza–Klein tower manifest, which was done in Giddings et al. (2000). This is not done here: we are interested in the full amplitude, and separating out the zero mode and Kaluza–Klein contributions doesn't help.

This is a bare Lagrangian, so $V(Q)$ is the classical tree-level potential for the quintessence. The brane metric $h_{ab}$ is just Minkowski space plus the perturbation $e_{ij}$, so $\psi$ (and $Q$) couple to gravity via the invariant volume element,

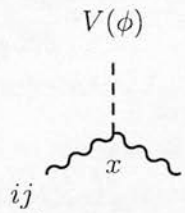
$$\sqrt{-\det h} \simeq 1 + \frac{e}{2} - \frac{e_{ij}e^{ij}}{2} + \frac{e^2}{8}. \tag{6.5.2}$$

This is evaluated using the same technique as (6.4.10). In particular, there is graviton–fermion vertex,




$$= -\mathrm{i}m\left(\frac{\delta_{ij}\delta_{rs}}{8} - \frac{\delta_{i(r}\delta_{s)j}}{4}\right)\varepsilon_B{}^A\delta_{\mathrm{D}}(z-R) \tag{6.5.3}$$

and a graviton–quintessence vertex,




$$= -\mathrm{i}V(Q)\left(\frac{\delta_{ij}\delta_{rs}}{8} - \frac{\delta_{i(r}\delta_{s)j}}{4}\right)\delta_{\mathrm{D}}(z-R) \tag{6.5.4}$$

We are ignoring any coupling between $\psi$ or $Q$ and $e_{ij}$ via the spin connexion. In principle these effects should be taken account, but they can reasonably be expected to give couplings at similar orders of magnitude to the couplings we are going to calculate. The procedure for translating perturbative path integrals into Feynman diagrams and Feynman integrals was traced in Chapter 2. In these diagrams, fermions are solid lines whereas gravitons are indicated by wavy lines and quintessence by dashes. The vertices are written in configuration space and are taken to occur at a spacetime point $x$, indicated on the diagram. Index pairs $ij$, $rs$ on graviton lines refer to the $\mathfrak{so}(3)$ index structure.

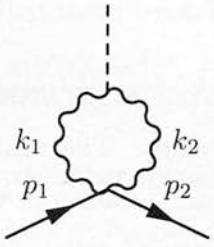
Fermions entering the diagram at a point $x$ with 4-momentum $p$ carry a coefficient function $(2\pi)^{-3/2}u(p)\mathrm{e}^{-\mathrm{i}p\cdot x}$, where $\cdot$ denotes the flat Euclidean inner product on the brane.[8] Fermions leaving the diagram from a point $x$ with momentum $p$ carry the conjugate coefficient function (Weinberg, 1994), $(2\pi)^{-3/2}\bar{u}(p)\mathrm{e}^{\mathrm{i}p\cdot x}$, where $\bar{u}$ denotes the Dirac conjugate of the spinor $u$, and spinor indices have temporarily been suppressed for clarity. Our conventions for spinors and $\gamma$-matrices match Weinberg (1994). Quintessence particles entering

[8] In principle, we ought to be using the spacetime inner product built from the metric $g_{ab}$. However this is not necessary provided one restricts attention to the brane, where the metric is flat.

or leaving the diagram carry $(2\pi)^{-3/2}\mathrm{e}^{-\mathrm{i}p\cdot x}$ or $(2\pi)^{-3/2}\mathrm{e}^{\mathrm{i}p\cdot x}$, respectively. To find amplitudes one integrates over the coordinates of all interaction points, including the appropriate spacetime volume measure $\sqrt{-\det g}$, which is unity on the brane.

## 6.6. Gravitational coupling of quintessence

**6.6.1. Loop diagram.** The lowest order diagram with only internal graviton lines which contributes to the effective quintessence mass is




Applying the Feynman rules, the amplitude for this process is obtained by multiplying the appropriate vertex factors with propagators to connect the initial and final vertices, and coefficient functions which carry momenta in and out of the diagram,

$$\Gamma_L = \frac{(-\mathrm{i})^4\kappa_5^4}{(2\pi)^{9/2}}\bar{u}(p_2)u(p_1)\int \mathrm{d}^5x\,\mathrm{d}^5y\,\mathrm{e}^{-\mathrm{i}p_1\cdot x+\mathrm{i}p_2\cdot x-\mathrm{i}p_0\cdot y}m\left(\frac{\delta_{ij}\delta_{pq}}{8}-\frac{\delta_{i(p}\delta_{q)j}}{4}\right)\delta_{\mathrm{D}}(z_x-R)$$
$$V(Q_{\mathrm{cl}})\left(\frac{\delta_{rs}\delta_{mn}}{8}-\frac{\delta_{r(m}\delta_{n)s}}{4}\right)\delta_{\mathrm{D}}(z_y-R)G^{ijrs}(x,y)G^{pqmn}(x,y), \tag{6.6.1}$$

where we have introduced a fictitious momentum $p_0$ which is carried away by the quintessence. The $\delta$-functions immediately restrict the support of the integrand to the brane at both vertices.[9] Expanding the definition of the propagators $G$, this can be written

$$\Gamma_L = \frac{1}{(2\pi)^{9/2}}\frac{16\kappa_5^2\ell^2}{R^2}\bar{u}(p_2)u(p_1)mV(Q_{\mathrm{cl}})\mathsf{G}\int\frac{\mathrm{d}^4x\,\mathrm{d}^4y\,\mathrm{d}^4k\,\mathrm{d}^4k'}{(2\pi)^8}\frac{1}{kk'}\frac{K_2(kR)}{K_1(kR)}\frac{K_2(k'R)}{K_1(k'R)}, \tag{6.6.2}$$

where $\mathsf{G}$ is a group-theory factor,

$$\mathsf{G}=\left(\frac{\delta_{ij}\delta_{pq}}{8}-\frac{\delta_{i(p}\delta_{q)j}}{4}\right)\left(\frac{\delta_{rs}\delta_{mn}}{8}-\frac{\delta_{r(m}\delta_{n)s}}{4}\right)\left(\delta^{i(r}\delta^{s)j}-\frac{1}{2}\delta^{ij}\delta^{rs}\right)\left(\delta^{p(m}\delta^{n)q}-\frac{1}{2}\delta^{pq}\delta^{mn}\right)=\frac{5}{8}. \tag{6.6.3}$$

[9] This is a considerable advantage of retaining the Feynman rules in coordinate form. In any case, the momentum space formulation is a little unusual, because there is no conserved Noether charge arising from translation invariance in the bulk which would correspond to a conserved bulk momentum. For this reason, loop integrals naturally involve circulating 4-momenta and not 5-momenta as one would naïvely expect.

Integrating out the coordinate functions to produce a momentum conservation $\delta$-functions at each vertex and simplifying numerical coefficients means that this can be reduced to a somewhat simpler expression,

$$\Gamma_L = \frac{10}{(2\pi)^{9/2}} \frac{mV(Q_{\rm cl})\kappa_5^2 \ell^2}{R^2} \bar{u}(p_2)u(p_1) \int \mathrm{d}^4k\, \mathrm{d}^4k'\, \frac{1}{kk'} \frac{K_2(kR)}{K_1(kR)} \frac{K_2(k'R)}{K_1(k'R)} \delta_{\rm D}(p_1 - p_2 - k - k')\delta_{\rm D}(k + k' - p_0). \tag{6.6.4}$$

The presence of Macdonald functions $K_\nu(z)$ makes this expression rather difficult to handle for arbitrary momenta $p_1$ and $p_2$. The analysis is much expedited by adopting an approximation in which all external 3-momenta vanish, so that $\mathbf{p}_1 = \mathbf{p}_2 = \mathbf{p}_0 = \mathbf{0}$, which is really just the non-relativistic approximation. This choice matches the analysis of Doran and Jäckel (2003). Additionally, with our conventions, at zero momentum the product of the coefficient functions $u(\mathbf{0})$, $\bar{u}(\mathbf{0})$, when summed over external spin states, is (Weinberg, 1994, Vol I, p. 222)

$$N^A{}_{\dot{A}} = \sum_{\sigma\in\{\uparrow,\downarrow\}} u^A(\mathbf{0},\sigma)\bar{u}_{\dot{A}}(\mathbf{0},\sigma) = \frac{1 + \mathrm{i}\gamma_0}{2}, \tag{6.6.5}$$

so $\bar{u}(\mathbf{0})u(\mathbf{0}) = \mathrm{Tr}\, N = 2$. This can also be shown just be evaluating the spin sum directly, using the result that the coefficient functions take the form

$$u(\mathbf{0},\uparrow) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad u(\mathbf{0},\downarrow) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \tag{6.6.6}$$

and the conjugation matrix $\beta = \mathrm{i}\gamma_0$,[10] where, in block diagonal form,

$$\gamma_0 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}. \tag{6.6.7}$$

[10] This is a flat space specialization of the more general result $\bar{u} = \mathrm{i}t^a\gamma_a u^\dagger$ for the Dirac adjoint, where $t^a$ is a unit timelike Killing vector.

Thus the amplitude, including external fermions in all spin states, is

$$\sum_{\sigma\in\{\uparrow,\downarrow\}}\Gamma_L=\frac{20}{(2\pi)^{9/2}}\frac{mV(Q_{\rm cl})\kappa_5^4\ell^2}{R^2}\delta_{\rm D}(p_1-p_2-p_0)\int \mathrm{d}^4k\frac{1}{k}\frac{K_2(kR)}{K_1(kR)}\,\frac{1}{k'}\frac{K_2(k'R)}{K_1(k'R)}\bigg|_{k'=p_1-p_2-k}. \tag{6.6.8}$$

The overall momentum conservation $\delta$-function can be discarded in this non-relativistic approximation. Because $p_1$ and $p_2$ are on-shell momentum 4-vectors for the same species of particle, the loop momenta are related via $k=k'$. After Wick rotating to Euclidean signature, one is left with the somewhat more tractable integral

$$\sum_{\sigma\in\{\uparrow,\downarrow\}}\Gamma_L=\frac{40\pi^2\mathrm{i}}{(2\pi)^{9/2}}\frac{mV(Q_{\rm cl})\kappa_5^4\ell^2}{R^2}\int_\mu^\Lambda \mathrm{d}k\,k\left(\frac{K_2(kR)}{K_1(kR)}\right)^2, \tag{6.6.9}$$

where we have explicitly written in an upper ultra-violet cut-off at Euclidean momenta $k\sim\Lambda$ and a lower infra-red cut-off at $k\sim\mu$. Although this regularization is extremely simple, it is easy to apply in the present context and moreover lets us assess the sensitivity of the amplitude $\Gamma$ to the cut-off. In four dimensions, $\Lambda$ will be of order the Planck scale, whereas in the brane world $\Lambda$ could (conceiveably) be as low as a TeV.

The Macdonald functions $K_\nu$ have asymptotics governed by

$$K_\nu(z)\overset{z\to0}{\longrightarrow}\frac{\Gamma(\nu)}{2}\left(\frac{z}{2}\right)^{-\nu}\quad\text{and}\quad K_\nu(z)\overset{z\to\infty}{\longrightarrow}\sqrt{\frac{\pi}{2z}}\mathrm{e}^{-z}. \tag{6.6.10}$$

In the infra-red, aside from numerical factors, the ratio $K_2(kR)/K_1(kR)$ behaves as a function of $k$ like $k^{-1}$. Combining this behaviour with the factor of $k^{-1}$ already present in (6.6.9), it can be seen that any infra-red divergence ought to be the same as in four dimensions. However, this is not the same for the upper cut-off $\Lambda$: the large-$z$ asymptotics of $K_\nu(z)$ changes the divergent behaviour in the ultra-violet. To make an estimate of the rough magnitude of (6.6.9), we write $\int_\mu^\Lambda=\int_\mu^{1/R}+\int_{1/R}^\Lambda$ and approximate the integrand using its asymptotic form in both regions, after changing variable to $z=kR$,

$$\int_\mu^\Lambda\left(\frac{K_2(kR)}{K_1(kR)}\right)^2\approx\frac{4}{R^2}\int_{\mu R}^1\frac{\mathrm{d}z}{z}+\frac{1}{R^2}\int_1^{\Lambda R}z\,\mathrm{d}z\approx-\frac{4}{R^2}\ln\mu R+\frac{1}{2}\Lambda^2. \tag{6.6.11}$$

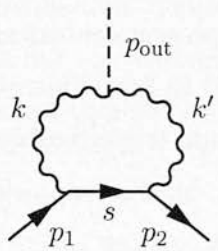
(We will carry out an approximation similar to this one in rather more detail in the next section. The skeptical reader is invited to peruse the more complicated calculation presented there before returning to verify that this approximation is correct.) We have discarded a term of order $O(1/R^2)$, which is a good approximation provided $\Lambda^2\gg1/R^2$.[11]

[11] Do not confuse the ultra-violet cut-off $\Lambda_{\rm UV}$ with the AdS radius $\ell^{-1}=\sqrt{\Lambda_{\rm AdS}/6}$. On-shell, where $R=\ell$, the condition for our approximation to be good is roughly that $\Lambda_{\rm UV}^2>\Lambda_{\rm AdS}$.

For an extra dimension of order 1 mm, $R^{-1} \sim 1.97 \times 10^{-4}\,\mathrm{eV}$, so this condition should be abundantly satisfied. This approximation lets us pick out the leading-order divergence in the ultra-violet and the infra-red as $\Lambda$ and $\mu$ are removed.

The ultra-violet divergence is logarithmic in the four-dimensional case (Doran and Jäckel, 2003) and here is modified to become quadratic. It is natural to interpret this modification as owing to interactions with the Kaluza–Klein tower. Despite this, the induced coupling remains proportional to the classical quintessence potential $V(Q)$, so this correction term does not destroy properties of the classical dynamics. This is entirely analogous to the situation in four dimensions.

**6.6.2. Triangle diagram.** The second important diagram contains an internal fermion line. It is




The corresponding amplitude satisfies

$$\Gamma_T = \frac{(-\mathrm{i})^6}{(2\pi)^{9/2}}\kappa_5^4 m^2 V(Q_{\mathrm{cl}})\bar{u}_{\dot{D}}(p_2)\varepsilon_A{}^{\dot{B}}\varepsilon_C{}^{\dot{D}}u^A(p_1)\frac{\delta_{ij}}{2}\frac{\delta_{rs}}{2}\left(\frac{\delta_{mn}\delta_{pq}}{8} - \frac{\delta_{m(p}\delta_{q)n}}{4}\right)$$
$$\int \mathrm{d}^5x\,\mathrm{d}^5y\,\mathrm{d}^5z\,\mathrm{e}^{-\mathrm{i}p_1\cdot x+\mathrm{i}p_2\cdot y-\mathrm{i}p_0\cdot z}\delta_{\mathrm{D}}(z_x-R)\delta_{\mathrm{D}}(z_y-R)\delta_{\mathrm{D}}(z_z-R)G^{ijmn}(x,z)G^{rspq}(y,z)$$
$$\int \frac{\mathrm{d}^4s}{(2\pi)^4}\frac{[-\mathrm{i}\not{s}+m]_{\dot{B}}{}^C}{s^2+m^2-\mathrm{i}\varepsilon}\mathrm{e}^{-\mathrm{i}s\cdot(x-y)} \tag{6.6.12}$$

By careful reduction, one arrives at the following simplification

$$\Gamma_T = -\frac{1}{(2\pi)^{9/2}}\frac{16\kappa_5^4\ell^2}{R^2}m^2V(Q_{\mathrm{cl}})\mathsf{G}\int\frac{\mathrm{d}^4k\,\mathrm{d}^4k'\,\mathrm{d}^4s}{(2\pi)^{12}}\,\delta_{\mathrm{D}}(p_1-s-k)\delta_{\mathrm{D}}(s-p_2-k')\delta_{\mathrm{D}}(k+k'-p_0)$$
$$\frac{\bar{u}(p_2)[-\mathrm{i}\not{s}+m]u(p_1)}{s^2+m^2-\mathrm{i}\varepsilon}\frac{1}{k}\frac{K_2(kR)}{K_1(kR)}\frac{1}{k'}\frac{K_2(k'R)}{K_1(k'R)}, \tag{6.6.13}$$

where $\mathsf{G}$ is the appropriate group-theory factor,

$$\mathsf{G} = \frac{\delta_{ij}}{2}\frac{\delta_{rs}}{2}\left(\frac{\delta_{mn}\delta_{pq}}{8} - \frac{\delta_{m(p}\delta_{q)n}}{4}\right)\left(\delta^{i(m}\delta^{n)j} - \frac{1}{2}\delta^{ij}\delta^{mn}\right)\left(\delta^{r(p}\delta^{q)s} - \frac{1}{2}\delta^{rs}\delta^{pq}\right) = \frac{1}{4}. \tag{6.6.14}$$

Integrating out the two $\delta$-functions leaves an overall momentum conservation $\delta$-function and a remaining loop integral,

$$\Gamma_T = -\frac{1}{(2\pi)^{9/2}} \frac{4\kappa_5^4 \ell^2}{R^2} m^2 V(Q_{\rm cl}) \delta_{\rm D}(p_1 - p_2 - p_0) \int \mathrm{d}^4 k \, \frac{\bar{u}(p_2)[-\mathrm{i}(\not{p}_1 - \not{k}) + m]u(p_1)}{(p_1 - k)^2 + m^2 - \mathrm{i}\varepsilon} \frac{1}{k} \frac{K_2(kR)}{K_1(kR)} \frac{1}{k'} \frac{K_2(k'R)}{K_1(k'R)} \Big|_{k'=p_1-p_2-k} \tag{6.6.15}$$

Since $u(p)$ is a momentum eigenstate, the Dirac equation says

$$-\mathrm{i}\not{\partial} u(p) = -\mathrm{i}\not{p} u(p) = m u(p). \tag{6.6.16}$$

This allows the numerator of the fermion propagator to be simpified, viz.

$$\bar{u}(p_2)[-\mathrm{i}(\not{p}_1 - \not{k}) + m]u(p_1) = \bar{u}(p_2)[\mathrm{i}\not{k} + 2m]u(p_1). \tag{6.6.17}$$

In the non-relativistic approximation the external momenta all vanish, so since $p_1$ and $p_2$ are on-shell $k = k'$ and the integral simplifies in the same way as the previous diagram. The remaining complication is

$$\mathrm{i}k_a \bar{u}(\mathbf{0})\gamma^a u(\mathbf{0}) \approx \mathrm{i}k_0 \bar{u}(\mathbf{0})\gamma^0 u(\mathbf{0}) = k_0. \tag{6.6.18}$$

(See Weinberg (1994, Vol I, p. 480).) When summed over external spins $\uparrow$, $\downarrow$ an extra factor of two arises, so the numerator reduces to $2k_0 + 4m$ before Euclidean continuation. At this point one would ordinarily complete the square in the denominator and drop terms which are odd in $k_a$, because the integral ought to be rotationally invariant. However, this procedure is inconvenient in the present case, since we wish to keep the argument of the Macdonald functions as simple as possible in order to exploit their asymptotics (Eq. (6.6.10)) for the purposes of making an approximate evaluation of the amplitude.

When carrying out the Wick rotation that results in the analytic continuation to Euclidean space, one must take care to keep track of factors of i which may accompany isolated components of $k_a$, such as $k_0$, and also the location of any poles of the integrand in the complex $k_0$-plane: carelessness with factors of i leads to integrals with obvious pathologies, which converge only for some values of $m$ and $R$, or not at all. This procedure is rather more complicated than the corresponding passage in the case of the loop diagram, which was isotropic, involving no angular integrations and only scalar quantities built out

of the $SO(1,3)$ or $SO(4)$ invariant $k$. The rate to be calculated is, omitting the overall momentum conservation $\delta$-function,

$$\sum_{\sigma\in\{\downarrow,\uparrow\}} \Gamma_T = -\frac{1}{(2\pi)^{9/2}}\frac{8\kappa_5^4\ell^2}{R^2}m^2V(Q_{\rm cl})\int \mathrm{d}^4k\,\frac{k_0+2m}{p_1^2-2p_1\cdot k+k^2+2m^2-\mathrm{i}\varepsilon}\frac{1}{k^2}\left(\frac{K_2(kR)}{K_1(kR)}\right)^2. \tag{6.6.19}$$

Consider the first factor in the integrand, which is the non-isotropic piece. The momentum $p_1$ is on-shell, so $p_1^2=-m^2$, which removes the other $m^2$ term from the denominator. In addition $p_1\cdot k$ must be $-mk_0$, so the denominator in this factor is

$$\frac{1}{k^2+2mk_0-\mathrm{i}\varepsilon} = \frac{1}{-k_0^2+2\mathbf{k}^2+2mk_0-\mathrm{i}\varepsilon}. \tag{6.6.20}$$

This has poles in the complex $k_0$-plane at

$$(k_0+m)^2=\mathbf{k}^2+m^2-\mathrm{i}\varepsilon \quad \text{so,} \quad k_0=\pm(\mathbf{k}^2+m^2)^{1/2}(1-\mathrm{i}\varepsilon). \tag{6.6.21}$$

These poles sit below the positive real axis and above the negative real axis, so there is no obstruction to rotation of the contour of integration through $\pi/2$ anticlockwise, provided that the isotropic factors contribute no troublesome singularities. The continuation prescription for the factor $k^{-2}$ is the same as for the fermion propagator, and it can be shown that the Macdonald functions are entire in the complex plane except for a singularity at the origin and an essential singularity at infinity.[12] Thus the standard Wick rotation prescription remains valid; this result was implicitly assumed when discussing the loop diagram. This rotation of the contour is effected by substituting $k_0\mapsto \mathrm{i}k_4$, where $k_4$ is integrated from $-\infty$ to $\infty$.[13] Thus,

$$\sum_{\sigma\in\{\downarrow,\uparrow\}} \Gamma_T = -\frac{\mathrm{i}}{(2\pi)^{9/2}}\frac{8\kappa_5^4\ell^2}{R^2}m^2V(Q_{\rm cl})\int \mathrm{d}^4k_{\rm E}\,\frac{\mathrm{i}k_4+2m}{k^2+2\mathrm{i}mk_4}\frac{1}{k^2}\left(\frac{K_2(kR)}{K_1(kR)}\right)^2. \tag{6.6.23}$$

The Euclidean volume element $\mathrm{d}^4k_{\rm E}$ can be expanded according to the usual Jacobian rule,

$$\mathrm{d}^4k_{\rm E}=k^3\,\mathrm{d}k\,\sin^2\theta\,\mathrm{d}\theta\,\sin\phi\,\mathrm{d}\phi\,\mathrm{d}\varphi = 4\pi k^3\,\mathrm{d}k\,\sin^2\theta \quad \text{after integrating out } \phi,\,\varphi, \tag{6.6.24}$$

[12] There are many equivalent ways to see that this is true, but a simple approach is just use the integral representation,

$$K_\nu(z)=\sqrt{\frac{\pi}{2z}}\frac{\mathrm{e}^{-z}}{\Gamma(\nu+1/2)}\int_0^\infty \mathrm{e}^{-t}t^{\nu-1/2}\left(1-\frac{t}{2z}\right)^{n-1/2}\mathrm{d}t. \tag{6.6.22}$$

[13] This is because the contour has been rotated anticlockwise. Had we rotated the real axis onto the imaginary axis clockwise, we should be replacing $k_0\mapsto \mathrm{i}k_4$, but integrating $k_4$ from $\infty$ to $-\infty$.

and we can choose the coordinate axes so that $k_4 = k\cos\theta$. It may appear that the amplitude resulting from this integral will be complex, which would be undesirable, but rotational invariance kills the imaginary part. Separating the integral into isotropic and angular-dependent parts, we have

$$\sum_{\sigma\in\{\downarrow,\uparrow\}} \Gamma_T = -\frac{\mathrm{i}}{(2\pi)^{9/2}} \frac{32\pi\kappa_5^4\ell^2}{R^2} m^2 V(Q_{\mathrm{cl}}) \int_\mu^\Lambda \mathrm{d}k \left(\frac{K_2(kR)}{K_1(kR)}\right)^2 I_\theta, \tag{6.6.25}$$

where

$$I_\theta = \int_0^\pi \mathrm{d}\theta\ \sin^2\theta \frac{\mathrm{i}k\cos\theta + 2m}{k + 2\mathrm{i}m\cos\theta} = \int_0^\pi \mathrm{d}\theta\ \sin^2\theta \frac{2mk(1+\cos^2\theta) + \mathrm{i}(k^2-4m)\cos\theta}{k^2 + 4m^2\cos^2\theta}. \tag{6.6.26}$$

Since $\cos\theta$ is odd on $[0,\pi]$ but the remainder of the integrand is even, the imaginary part vanishes. Making the replacement $z = kR$ leads to the integral we shall actually calculate,

$$\sum_{\sigma\in\{\downarrow,\uparrow\}} \Gamma_T = -\frac{\mathrm{i}}{(2\pi)^{9/2}} \frac{64\pi\kappa_5^4\ell^2}{R^2} m^3 V(Q_{\mathrm{cl}}) \int_{\mu R}^{\Lambda R} z\,\mathrm{d}z \left(\frac{K_2(z)}{K_1(z)}\right)^2 \frac{\sin^2\theta(1+\cos^2\theta)}{z^2 + 4m^2R^2\cos^2\theta}. \tag{6.6.27}$$

The process of making an estimate of the infra-red and ultra-violet divergences of this integral is somewhat facilitated by actually carrying out the angular part,

$$\int_0^\pi \mathrm{d}\theta\ \frac{\sin^2\theta(1+\cos^2\theta)}{z^2 + 4m^2R^2\cos^2\theta} = \frac{\pi}{z^2 + z\sqrt{4m^2R^2+z^2}} + \frac{\pi}{8m^2R^2 + 4z^2 + 4z\sqrt{4m^2R^2+z^2}}, \tag{6.6.28}$$

which can be verified without the necessity for laborious calculation using Mathematica, Maxima or any other symbolic algebra software. In the ultra-violet, $K_2(z)/K_1(z) \to 1$, so the divergence is like

$$\int^{\Lambda R} z\,\mathrm{d}z \left(\frac{\pi}{2z^2} + \frac{\pi}{8z^2}\right) = \int^{\Lambda R} \frac{5\pi}{8z}\,\mathrm{d}z = \frac{5\pi}{8}\ln\Lambda R, \tag{6.6.29}$$

whereas $K_2(z)/K_1(z) \to 2/z$ as $z \to 0$, so in the infra-red one has

$$\int_{\mu R} z\,\mathrm{d}z\ \frac{4}{z^2}\frac{\pi}{2mRz} = \frac{2\pi}{mR}\int_{\mu R}\frac{\mathrm{d}z}{z^2} = \frac{2\pi}{3mR}\frac{1}{(\mu R)^3}. \tag{6.6.30}$$

Therefore, the leading divergence in the amplitude has the form

$$\sum_{\sigma\in\{\downarrow,\uparrow\}} \Gamma_T = -\frac{\mathrm{i}}{(2\pi)^{9/2}} \frac{64\pi^2\kappa_5^4\ell^2}{R^2} m^3 V(Q_{\mathrm{cl}}) \left(\frac{5}{8}\ln\Lambda R + \frac{2}{3mR}\frac{1}{(\mu R)^3}\right), \tag{6.6.31}$$

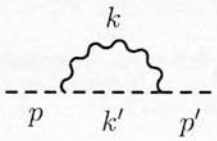
and, as before terms of order $1/R^2$ have been discarded. One can verify that the introduction of the fermion propagator has contributed an extra momentum factor of $1/p^2$ to the

integrand, since the linear piece in the fermion numerator vanished by rotational invariance. All the same, this amplitude is proportional to $V(Q_{\rm cl})$ and therefore can be absorbed into a redefinition of the overall scale of the classical potential. Just like the loop diagram, the classical potential is not destroyed.

At this point it is rather clear that this result is general. Since the graviton couples to quintessence through a vertex factor, which does not change when jumping from four-dimensional cosmology to the braneworld, the result is the same as Doran and Jäckel (2003), even though the *character* of the divergences has been modified. From the point of view of particle phenomenology, this modification of the divergences is more interesting result. As a final observation, we record that the coupling of quintessence to gravity via the spin-connexion would typically be expected to modify the quintessence potential. At the time of writing, no calculations of this effect exist.

### 6.7. Vacuum polarization and the quintessence mass

As a final test, we suppose that the quintessence particle has some bare mass $M_Q$ and calculate the shift induced by the contribution of gravitons circulating in a loop.




This is a contribution to the quintessence self-energy, or vacuum polarization, of the sort which was discussed fairly extensively in Section 2.3.3, leading to the expression (2.3.31) for the dressed momentum space propagator. In our units and the present notation, this is is

$$\langle Q(x_1)Q(x_2)\rangle = -\mathrm{i}\int \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{k^2+M_Q^2-\Pi^*(k)}\mathrm{e}^{-\mathrm{i}k\cdot(x_1-x_2)}, \tag{6.7.1}$$

where the vacuum polarization is written $\Pi^*$, as conventional for scalar particles (the notation $\Sigma$ is usually reserved for fermions) and $\hbar = 1$. We assume that the quintessence potential is purely a mass term, so the theory is still trivial and all integrals are Gaussian,

$$S_Q = \frac{1}{2}\int \mathrm{d}^4x\ \partial_a Q\partial^a Q + M_Q^2Q^2. \tag{6.7.2}$$

The coupling to gravity we are going to consider is via the $\mathrm{Tr}\, e/2$ term, leading to the vertex shown above. With propagators for quintessence particles appropriately stripped

off, the momentum-space vacuum polarization is

$$\mathrm{i}(2\pi)^4\Pi^*(p) = (-\mathrm{i})^4\frac{\kappa_5^2\ell}{R}\frac{M_Q^4}{4}\mathsf{G}\int \mathrm{d}^5x\,\mathrm{d}^5y\;\delta_{\mathrm{D}}(z_x-R)\delta_{\mathrm{D}}(z_y-R)\mathrm{e}^{\mathrm{i}x\cdot p_1}\mathrm{e}^{-\mathrm{i}y\cdot p_2}\mathrm{e}^{\mathrm{i}k\cdot(x-y)}\mathrm{e}^{\mathrm{i}s\cdot(x-y)}$$
$$\int\frac{\mathrm{d}^4k\,\mathrm{d}^4s}{(2\pi)^8}\frac{1}{k}\frac{K_2(kR)}{K_1(kR)}\frac{1}{s^2+m^2-\mathrm{i}\varepsilon}, \tag{6.7.3}$$

where momentum $p_1$ is carried into the diagram and momentum $p_2$ is carried out. The group-theory factor is

$$\mathsf{G} = \delta_{ij}\delta_{mn}(\delta^{i(m}\delta^{n)j} - \frac{1}{2}\delta^{ij}\delta^{mn}) = -4, \tag{6.7.4}$$

so after integrating out the $\delta_{\mathrm{D}}$-functions and the coordinate space integrals, we are left with

$$\mathrm{i}(2\pi)^4\Pi^*(p) = -\frac{\kappa_5^2\ell}{R}M_Q^4\int\mathrm{d}^4k\;\frac{1}{k}\frac{K_2(kR)}{K_1(kR)}\frac{1}{(p-k)^2+m^2}. \tag{6.7.5}$$

in which the overall momentum conservation $\delta_{\mathrm{D}}$-function, which was only present for notational purposes, has been omitted and we have set $p_1 = p_2 = p$. In principle, one should leave the $p$ dependence intact, as we use a relation like (2.3.40) to fix the renormalized mass: essentially, this is equivalent to demanding that the dressed quintessence propagator has a pole at the location of the renormalized mass $\tilde{M}_Q$. In the present case, however, that calculation is long, messy and algebraic and does not lead to any useful insights. A better strategy, and certainly a more transparent calculation, is to note that the quintessence mass $M_Q$ is very small, and quantum effects cannot renormalize it much. In that case, to be self-consistent, it should be a passable approximation to fix

$$p^2 = -\tilde{M}_Q^2 \approx M_Q \tag{6.7.6}$$

so evaluating the vacuum polarization at zero momentum transfer with $p$ on-shell should give a passable approximation to the mass shift. After carrying out this procedure, a Wick rotation to Euclidean signature, and dropping terms which vanish owing to rotational invariance, one is left with

$$\mathrm{i}(2\pi)^4\Pi^*(p) = -\frac{4\pi\mathrm{i}}{R^2}\kappa_5^2\ell M_Q^4\int_{\mu R}^{\Lambda R} z^2\,\mathrm{d}z\;\frac{K_2(kR)}{K_1(kR)}\frac{\sin^2\theta}{z^2+4m^2R^2\cos^2\theta}. \tag{6.7.7}$$

The angular integral is

$$\int_0^\pi\frac{\sin^2\theta}{z^2+4m^2R^2\cos^2\theta} = \frac{\pi}{z^2+z\sqrt{4m^2R^2+z^2}}, \tag{6.7.8}$$

Estimating the integral is the asymptotically large and small $z$ régimes, using the same ideas outlined in the previous section, shows that the leading divergence must have the form

$$\mathrm{i}(2\pi)^4\Pi^* \sim -\frac{2\pi^2\mathrm{i}\ell}{R}\kappa_5^2 M_Q^4\Lambda. \tag{6.7.9}$$

This is an entirely ultra-violet effect. There is no infra-red divergence. On-shell, where $R = \ell$, our estimate of the mass shift therefore reads

$$\delta m^2 \approx -\Pi^* \sim \frac{1}{\pi}\frac{\Lambda M_Q^4}{M_5^3}, \tag{6.7.10}$$

where $M_5$ is the five-dimensional Planck scale, and we have remembered that $\kappa_5^2 = 8\pi M_5^{-3}$, which contains an extra power of the Planck scale in comparison with the four-dimensional coupling. It is easy to verify that this expression is dimensionally correct. To find a numerical estimate, observe that in (4.5.4) it was shown that a generically sensible value for the mass of a scalar fields driving any inflationary epoch, such as the quintessence field $Q$ driving the late-time acceleration of the universe, is the epochal Hubble rate, or in the present case

$$M_Q \sim 2.1 \times 10^{-33} h\,\mathrm{eV} \tag{6.7.11}$$

where, as usual, $h$ is the experimental Hubble rate in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. If $M_5$ is of order a TeV, or $M_5 \sim 10^{12}\,\mathrm{eV}$, then

$$\delta m^2 \sim 10^{-108} h^4\,\mathrm{eV}^2. \tag{6.7.12}$$

There is a fairly straightforward comparison calculation which can be made in four dimensions. The vacuum polariation is

$$\mathrm{i}(2\pi)^4\Pi_4^*(p) = (-\mathrm{i})^4\frac{\kappa_4^2}{4}M_Q^4\mathsf{G}\int \mathrm{d}^4k\,\frac{1}{k^2+m^2}\frac{1}{(p-k)^2}, \tag{6.7.13}$$

where $\mathsf{G}$ is the same group-theory factor as in five dimensions. Collecting terms, translating to Euclidean space, and at zero momentum transfer, that is

$$\mathrm{i}(2\pi)^4\Pi_4^* = -4\pi\mathrm{i}\kappa_4^2 M_Q^4\int_\mu^\Lambda \frac{s\,\sin^2\theta}{s^2+4M_Q^2\cos^2\theta}. \tag{6.7.14}$$

This is infra-red safe, so a good estimate of the leading divergence is

$$\mathrm{i}(2\pi)^4\Pi_4^* \sim -2\pi^2\mathrm{i}\kappa_4^2 M_Q^4\ln\frac{\Lambda}{\mathrm{eV}}. \tag{6.7.15}$$

This translates to a mass shift of the order

$$\delta m^2 \sim \frac{2}{\pi}\frac{M_Q^4}{M_4^2}\ln\frac{\Lambda}{\mathrm{eV}}, \tag{6.7.16}$$

in which $M_4$ is the four-dimensional Planckmass, $M_4 \sim 10^{19}\,\mathrm{GeV}$. Numerically this gives an estimate $\delta m^2 \sim 10^{-187}h^4\,\mathrm{eV}^2$. The ratio is

$$\frac{\delta m^2(5\mathrm{D})}{\delta m^2(4\mathrm{D})} = \frac{1}{2}\Lambda_5\frac{M_4^2}{M_5^3}\ln\frac{\Lambda_4}{\mathrm{eV}}. \tag{6.7.17}$$

## 6.8. Summary

We have studied the constraints that are on TeV-scale quintessence models from a variety of sources. Non-renormalizable operators in the four-dimensional theory, arising from integrating out details of the higher-energy physics, will typically be important. Bearing this in mind, it is clear that the quintessence potential should really be calculated from the five-dimensional framework. Four-dimensional results cannot always be trusted. This follows from the fundamental mismatch between the scale $M \sim$ TeV which determines the scale at which non-renormalizable operators become important, and the vacuum expectation value $\langle Q\rangle$ of the quintessence field. $\langle Q\rangle$ is typically of the order of the four dimensional Planck mass in tracker models. Perturbation theory in $Q/M_5$ fails spectacularly.

In contrast, the gravitational coupling of quintessence to fermionic matter in cosmologies of the Randall–Sundrum type does *not* yield significant constraints, because the controlling physics does not change in moving between frameworks, provided field theory is applicable. We find that one-loop effects introduce quantum corrections in the effective potential which are proportional to the classical potential, and can just be absorbed into an overall renormalization of scale. This is exactly the same as in four dimensions. The vertices of the diagram generate the couplings, not propagators, and the propagator is the essential modification of the Feynman rules in moving between conventional cosmology and the brane world. In fact, this is part of a much more general result, that whatever kind of low-energy gravity appears, provided it can be described by a field theory the quintessence potential will not receive destructive renormalizations from 'minimal coupling'. As we indicated in the text, coupling via the spin connexion generically *will* destroy the quintessence potential, but the level of the effect is not known.

We have also computed the lowest-order contribution from graviton loops to the vacuum polarization of quintessence. In this case one must make a numerical estimate, and one

finds that the brane universe induces an approximate mass shift much larger than in four dimensions. The shift is dependent on the ultra-violet cut-off, and scales with the expected ratio of four- and five-dimensional Planck scales. From the point of view of an observer on the brane, the Planck scale part is just a reflection of the different strength of gravitational couplings in four and five dimensions, and does not depend on the fact that one is in a warped compactification or that there is a tower of Kaluza–Klein modes in the theory. The other piece controlling the ratio, which has the form $\Lambda_5 \ln \Lambda_4$, is dependent on the exact relationship between four- and five-dimensional gravity, and *is* sensitive to the nature and chacteristics of the compactification. In this case, it is seeing the effect of the non-trivial conformal field theory which controls the brane.

CHAPTER 7

# Bulk quantum fields, perturbation theory, and breathing orbifolds

## 7.1. Introduction

The recent experimental results from the WMAP project (see Section 4.9; Spergel et al. (2003)), have lent rather strong support to the commonly accepted view that the observed homogeneity, isotropy and large-scale structure of the universe arises from an early period of accelerated expansion, dubbed inflation (see Chapter 4; Blau and Guth (1987); Guth (1981)). The general principles of inflation were set out in Section 4.5, where it was explained that the expansion is commonly supposed to be driven by a light scalar field, or inflaton, that violates the weak energy condition and dominates the energy density of the universe. During inflation all massless or sufficiently light[1] degrees of freedom are quantum mechanically excited (Section 4.6), and pick up a nearly scale invariant fluctuation. The characteristics of this fluctuation are controlled by the expansion rate, and hence, via the Friedmann equation, depend on the inflaton potential. In the later universe, after the inflaton decays and reheats the universe, the inflaton fluctuation is communicated to the curvature of spatial slices ((4.6.13); Mukhanov et al. (1992)). This process seeds primordial structure formation, and is observable today in the large-scale distribution of the galaxies (Hawkins et al., 2002) and fluctuations in the cosmic microwave background (CMB) (Hu and Sugiyama, 1995).

The curvature fluctuation need not be the only fossil from the early universe (cf. (5.7.1)). Since the graviton is massless, one would expect it to acquire a similar fluctuation in which small tensor perturbations would have been excited. Significantly, the subsequent evolution of these tensor perturbations differs from that of the inflaton, because they do not decay: the relative weakness of the gravitational coupling implies, in fact, that tensor perturbations would essentially not interact with other constituents of the universe

[1]This excitation process works for fields whose mass $m$ is less than $3H/2$, where $H$ is the Hubble rate during inflation.

on their journey towards us. Therefore such perturbations would almost certainly still be in their primordial state and could offer great insight into conditions and physics of the early universe. Tensor perturbations of this type are in principle observable today as a stochastic background of gravity waves, and could eventually be measured via their imprint in the polarization field of the CMB (Bond and Efstathiou, 1984; Kamionkowski, Kosowsky, and Stebbins, 1997; Polnarev, 1985; Zaldarriaga and Seljak, 1997) by experiments such as QUEST (Bowden et al., 2004). In the short term, it is more appropriate to hope for experimental conformation that gravity waves exist at all from gravity wave observatories such as LIGO or GEO (Abbott, 2003; Shoemaker, 2004). In particular, should the gravity wave power spectrum be observed in the near future, then in addition to the already observed amplitude $A_S^2$ and spectral index $n_S$ of the scalar spectrum, one would also have similar information $A_T^2$, $n_T$ available for tensor perturbations. Such extra information would be of great importance for cosmology, and cosmological parameter estimation (Liddle, 2004).

In the context of scalar field inflation, all details of the power spectrum, including amplitudes and spectral indices, are determined by properties of the scalar potential. In the slow-roll formalism this means they can be expressed in terms of the characteristic numbers $\varepsilon$ and $\eta$, at lowest order. This dependency on a single source implies that one might expect to find some relations between the various measurable quantities. For example, in the special case of the standard cosmology, one finds (Starobinsky, 1985), to lowest order in the slow-roll approximation (Liddle and Lyth, 2000),

$$n_T \approx -2\frac{A_T^2}{A_S^2}. \tag{7.1.1}$$

This is not an exact relationship. It is proved by expanding both sides in terms of the slow-roll parameters, and noticing that they agree to lowest order. We would like to stress that, in the context of scalar field inflation, whatever exact relation exists between observables is not known. Indeed, we do not even have an exact expression for $A_S^2$ or $A_T^2$; all that is known is a perturbation expansion in the slow-roll parameters. For this reason, if inflation was not of the slow-roll variety, then higher-order terms in the expansion could be important, and (7.1.1) might not apply.

Eq. (7.1.1) is at present only a theoretical prediction. However, it is one of only a handful of testable predictions made by the inflationary paradigm, and for this reason has the potential to be a powerful discriminant between competing models. Over and above

the general current evidence in favour of an inflationary-like epoch, an observation of this relation in the real universe would provide extremely strong support for a minimal scalar field model. On the other hand, more complex models weaken (7.1.1) to an inequality; this occurs in models containing isocurvature modes. Therefore observations of gravity waves at a lower level than predicted by (7.1.1) can be consistent with inflation, whereas observing an excess of primordial gravitational power would be a severe blow to the inflationary programme. (7.1.1) is of considerable observational importance, and we will discuss how it arises in more detail below.

One can calculate a similar equation that is exact to next-order in the slow-roll expansion (Lidsey et al., 1997). This next-order term does not preserve the functional form of (7.1.1); instead, one has

$$n_T = -2\frac{\Delta_T^2}{\Delta_S^2}\left[1 - \frac{\Delta_T^2}{\Delta_S^2} + (1 - n_S)\right]. \tag{7.1.2}$$

This is proved by expanding both sides in terms of the slow-roll parameters to next-order. Eq. (7.1.1) is informally called the inflationary consistency relation; Eq. (7.1.2) is sometimes known as the next-order consistency relation. We will return to this equation later, giving a brief derivation in Section 7.1.1 and considering first order perturbations in Sections 7.4–7.5. The analysis of circumstances in which (7.1.1) may *fail* to hold is the principal concern of this chapter.

Because (7.1.1) is a delicate prediction involving gravity and quantum field theory, one might imagine that it would easily break when translated to the braneworld models described in Chapter 5. One can calculate the modifications to the predictions for late-universe observables in the Kaloper–Linde model, which acts as an eternal de Sitter phase (Bridgman et al., 2002; Frolov and Kofman, 2002; Giudice et al., 2002; Gorbunov et al., 2001; Langlois et al., 2000; Maartens, Wands, Bassett, and Heard, 2000), but one discoverse a remarkable surprise. Although predictions for the tensor and scalar amplitudes and spectral indices are modified, as a result of their sensitivity to the behaviour of gravity in the large extra dimensions, the lowest-order consistency relation survives (Huey and Lidsey, 2001, 2002).

This is a non-trivial feature of the model, and at the time of writing there is no simple argument which demonstrates why it should be true. Because of its potential observational importance, this is both an immediate and pressing observational difficulty,

and a challenging theoretical puzzle. The continued appearance of (7.1.1) in the braneworld potentially jeopardizes the long-standing hope of observationally reconstructing the inflaton potential (Liddle and Taylor, 2002; Lidsey et al., 1997). An understanding of the origin of the braneworld degeneracy is essential to the reconstruction programme, for if such degeneracies apply to an open set of models then it may be difficult or impossible to place confidence in inflaton protentials reconstructed from minimal scenarios. One should notice, however, that this consistency relation is derived assuming that the bulk is empty. This need not be the case. If the bulk contains a scalar field, then there is typically a tachyonic instability (Frolov and Kofman, 2004). Requiring that this instability is stabilized can only be achieved in a régime where fluctuations of the bulk scalar field dominate the scalar perturbation as seen on the brane. In this case, the scalar perturbation spectrum does not coincide with the spectrum assumed above, and the consistency relation (7.1.1) is destroyed (Frolov and Kofman, 2004).

In this chapter, we clarify the circumstances under which one expects degeneracies between brane cosmology and conventional cosmology to persist. This programme is carried out by marginally perturbing the cosmology which gives rise to (7.1.1), and asking if the consistency relation is still satisfied in the perturbed cosmology. It is possible to solve both the gravitational and scalar field equations for the power spectra and spectral indices. This is fairly straightforward in the scalar case, but gravitational perturbations cannot be handled so easily and our technique requires a considerable extension of existing methods. We do not seek to provide a mechanism from which the perturbation may originate. In five dimensions one can appeal to possible brane–bulk interactions, but it is also possible to regard the cosmology as simply a model for a universe which is close to the de Sitter state, but does not exactly coincide with it.

This chapter is organized as follows. The four-dimensional amplitudes and spectral indices were derived in Section 4.6.2, and the braneworld theory was given in Section 5.7.2, principally as a canonical quantum field theory. In this chapter we are more interested in the path integral theory, since it is this approach which generalizes most easily to perturbed cosmologies. We briefly introduce the path integral (Section 7.2), and properly derive the power spectrum (5.7.60). In Section 7.1.1 we discuss the consistency relation in the unperturbed four- and five-dimensional cases, and show how it arises. In Sections 7.4–7.5 we calculate the effect of an arbitrary perturbation of the Hubble rate, $\delta H$, on the

power spectra of scalar and tensor quantities, and study the effect on the consistency relation. This is done in both the four- and five-dimensional cases. We begin with an exact de Sitter cosmology fixed by $H =$ constant, and assume it is still valid to treat the field fluctuations such as $\phi$ as free, massless fields propagating over the background. The two-point correlation functions can then be calculated. We demonstrate explicitly that the perturbation comes entirely from the matter sector on the brane, and does not involve dark radiation or other unknown physics impinging from the bulk which might reasonably be expected to trivially alter four-dimensional physics.

## 7.2. Braneworld power spectrum

Let $\phi$ be a free, massless scalar field. Its correlation functions are controlled by the functional integral (Section 2.3.1),

$$\langle \phi(x_1) \cdots \phi(x_n) \rangle = \int [\mathrm{d}\phi]\, \phi(x_1) \cdots \phi(x_n) \exp\left( -\frac{\mathrm{i}}{2} \int_{\mathcal{M}} \mathrm{d}v\, \phi \Box \phi \right), \tag{7.2.1}$$

where $[\mathrm{d}\phi]$ is the funtional measure, $\mathcal{M}$ is the background spacetime with metric $g_{ab}$ and invariant volume measure $\mathrm{d}v$, and we have chosen units in which $\hbar = 1$. The operator $\Box$ is defined by $\Box = \nabla^a \nabla_a$, where $\nabla_a$ is the covariant derivative compatible with $g_{ab}$. In particular, the two-point function satisfies $\langle \phi(x_1)\phi(x_2) \rangle = -\mathrm{i}\Box^{-1}(x_1, x_2)$ (Weinberg, 1994).

Now let $\mathcal{M}$ be four-dimensional de Sitter space. We choose local coordinates in which the metric takes the form (Hawking and Ellis, 1973)

$$\mathrm{d}s^2 = \frac{1}{H^2\tau^2} \left( -\mathrm{d}\tau^2 + \delta_{ij} \mathrm{d}x^i \mathrm{d}x^j \right). \tag{7.2.2}$$

This form of the metric with flat spatial slices is particularly common and convenient when discussing inflation. The infinite past corresponds to $\tau \to -\infty$. The calculation of the two-point function in de Sitter space is a minor modification of the argument presented in Section 4.6.2, where scalar field fluctuations were coupled to metric perturbations. The power spectrum satisfies $A_\phi^2 = (H/2\pi)^2$, leading to a matter amplitude (4.6.85) and spectral index (4.6.86).

Now consider the passage to the braneworld. The scalar power spectrum is trivial, and does not much affect the theory to be set out in this chapter, so we shall be brief. Let $\phi$ be a free massless scalar field propagating over the brane, $\Sigma$. Then the propagator for $\phi$ is still defined by (7.2.1) (with integration over spacetime $\mathcal{M}$ replaced by integration over the

slice $\Sigma$ which corresponds to our universe) and is exactly the same as the four-dimensional case, $A_\phi^2 = (H/2\pi)^2$. This is consequence of the observation that it does not matter how a de Sitter geometry arises when calculating the power spectrum of a scalar field (Wands et al., 2000).

The situation for gravitational perturbations is more complicated, and was first analysed by Langlois et al. (2000) for the case of a Kaloper–Linde type brane; see also Gorbunov et al. (2001) for a more detailed treatment. The simpler case of tensor perturbations around flat Randall–Sundrum branes was studied in detail by Giudice et al. (2002). There is an alternative formalism, based on the AdS/CFT correspondence (discussed in Section 5.4), in the special case that the brane carries a large $N$ CFT. This was first done by Nojiri and Odintsov (2000) and later by Hawking et al. (2000) (see also Nojiri, Odintsov, and Zerbini (2002)).

Langlois et al. (2000) worked in the Schrödinger picture. This approach was formalised in Section 5.7.1 where the canonical quantum theory was contructed. The field $e_{ij}$ is a small perturbation (as described by (5.7.1)) of the braneworld metric, and for the purposes of this chapter can be taken to be transverse and traceless, although in principle one should include Fadeev–Popov and ghost terms (Section 2.4.2). As before, we ignore the Kaluza–Klein vector and scalar particles, which do not contribute during inflation (Frolov and Kofman, 2002; Langlois et al., 2000). The two-point function for $e_{ij}$ satisfies

$$\langle e^{ij}(x_1)e_{rs}(x_2)\rangle = \int [\mathrm{d}e_{mn}]\, e^{ij}(x_1)e_{rs}(x_2) \exp\left[-\frac{\mathrm{i}}{8\kappa_5^2}\int_{\mathcal{M}} \mathrm{d}v\, e^{mn}\left(\frac{\Box_\parallel}{n^2}+\Box_\perp\right)e_{mn}\right], \tag{7.2.3}$$

where we have decomposed the five-dimensional braneworld scalar d'Alembertian, $\Box = \nabla^a\nabla_a = \Box_\parallel/n^2 + \Box_\perp$, into two terms $\Box_\parallel$ and $\Box_\perp$, defined by

$$\Box_\parallel = -\frac{\partial^2}{\partial t^2} - \left(3\frac{\dot{a}}{a} - \frac{\dot{n}}{n}\right)\frac{\partial}{\partial t} + \frac{n^2}{a^2}\triangle, \tag{7.2.4a}$$

$$\Box_\perp = \frac{\partial^2}{\partial y^2} + \left(3\frac{a'}{a} + \frac{n'}{n}\right)\frac{\partial}{\partial y}. \tag{7.2.4b}$$

Because (5.1.1) is not a product metric $\Box_\parallel$ and $\Box_\perp$ are not the on- and off-brane d'Alembertians, but in the important special case that the brane is endowed with a de Sitter geometry $\dot{H} = 0$ these operators separate (Langlois et al., 2000) – this is the Kaloper–Linde model. In this case $\Box_\perp$ is an honest Sturm–Liouville operator and one can write $E_{ij}$ as a sum over its eigenfunctions. The possibility that this unwrapping occurs can be understood in terms of

the bundle structures for braneworld compactifications outlined in Chapter 5, and allows us to re-express the path integral measure as a product of four-dimensional path integrals, essentially rewriting the whole theory as an effective four-dimensional theory with an infinite tower of increasingly massive fields as in the usual string compactification scenario. This strategy is key to the solubility of the de Sitter model. When considering the more general, perturbed theory later we will have to largely abandon this approach, although we retain some of its aspects.

Following Langlois et al. (2000) (and (5.7.19a)–(5.7.19b)) we define a set of weighted eigenfunctions, $\mathcal{E}_\alpha(y)$, of $\Box_\perp$ by

$$\Box_\perp \mathcal{E}_\alpha(y) = -\frac{\alpha^2}{n^2}\mathcal{E}_\alpha(y). \tag{7.2.5}$$

This is a reasonable Sturm–Liouville operator (cf. Appendix A), so the $\mathcal{E}_\alpha$ can be chosen to be orthonormal (cf. (5.7.31)–(5.7.33)),

$$2\int_0^{y_\mathrm{h}} n^2 \,\mathrm{d}y\, \mathcal{E}_\alpha^* \mathcal{E}_\beta = \delta_{\alpha\beta}, \tag{7.2.6}$$

provided the $\mathcal{E}_\alpha$ obey suitable boundary conditions at $y = 0$ and $y = y_\mathrm{h}$, essentially (5.7.32). We have added a factor 2 by hand in the normalization to take account of the other branch of the orbifold. The allowed boundary conditions for the $\mathcal{E}_\alpha$ (rather restrictive by comparison with the low-spin cases) have already been outlined (cf. (5.7.7)), and require that $\mathcal{E}'_\alpha = 0$ at the brane and $y = y_\mathrm{h}$. The spectrum then follows the general scheme enforced by the scaling argument (5.7.38) which gives the conformal dimension of the AdS/CFT bulk fields, and consists of a bound zero mode at $\alpha = 0$ and a continuum of massive modes for $\alpha > 3H/2$. The standard Sturm–Liouville argument (Section A.1.1.3; Kolmogorov and Fomin (1957); Riesz and Sz.-Nagy (1955)) guarantees that a field such as $e_{ij}$ can be expressed almost everywhere as an admixture of the $\mathcal{E}_\alpha$. Performing this decomposition for $e_{ij}$, expressing the path integral measure in the same terms, and integrating over the transverse dimension in the action gives, for coordinates $x_1$, $x_2$ on the brane,

$$\langle e^{ij}(x_1) e_{rs}(x_2)\rangle \simeq \prod_\alpha \mathcal{E}_\alpha^2(0) \int [\mathrm{d}e_\alpha^{mn}]\, e_\alpha^{ij}(x_1) e_{rs,\alpha}(x_2) \exp\left(-\frac{\mathrm{i}}{8\kappa_5^2}\int_\Sigma \mathrm{d}v\; e_{mn,\alpha}(\Box - \alpha^2) e_\alpha^{mn}\right), \tag{7.2.7}$$

where $e^{ij}(x,y) = \sum_\alpha e_\alpha^{ij}(x)\mathcal{E}_\alpha(y)$, $\mathrm{d}v$ is the volume measure on the de Sitter slice $\Sigma$, $\Box$ is the de Sitter Laplacian, and there are off-diagonal terms proportional to $\mathcal{E}_\alpha \mathcal{E}_{\alpha'}$ $(\alpha \neq \alpha')$ which we have neglected. These cross-modes can reasonably be expected to exhibit destructive

interference and should not acquire signficant correlation functions. Thus the field $e_{ij}$ behaves like a collection of four-dimensional Klein–Gordon fields in de Sitter space, with masses described by the allowed values of $\alpha$. At low energies, or during inflation, only the $\alpha = 0$ zero-mode will be excited, so since $\mathcal{E}_0$ is independent of $y$, one has

$$\langle e^{ij}(x_1)e_{rs}(x_2)\rangle = \mathcal{E}_0^2 \int [\mathrm{d}e_0^{mn}]\, e_0^{ij}(x_1)e_{rs,0}(x_2) \exp\left(-\frac{\mathrm{i}}{8\kappa_5^2}\int_\Sigma \mathrm{d}x\; e_{mn,0}\Box e_0^{mn}\right). \quad (7.2.8)$$

The higher modes have masses $\alpha > 3H/2$ which are heavy during inflation and are left in their vacuum states (Langlois et al., 2000). Therefore, the dominant contribution coming from the zero mode is no more than the standard result, with an extra factor of $\mathcal{E}_0^2$ in the normalization. This is still a free theory, so there is no obstruction to taking the coincidence limit $x_1 \to x_2$ as in four dimensions, and the power spectrum follows. Eliminating the five-dimensional coupling $\kappa_5^2$ in favour of its four-dimensional counterpart, gives a final result

$$A_T^2 = \frac{\kappa_4^2 F^2}{50}\frac{H^2}{\pi^2}. \quad (7.2.9)$$

where $\mathcal{E}_0^2 = \mu F^2$. The quantity $F$, defined in (5.7.55), expresses a renormalization of the amplitude $A_T^2$ in comparison with the four-dimensional result. This can be understood as a volume term arising from integrating out the extra dimension: the zero mode is a collective excitation which couples in the same way as the four-dimensional metric (Verlinde, 2000). This is the result (5.7.60) which was quoted for the tensor spectrum in the braneworld in Chapter 5.

The renormalization $F$ was derived in Chapter 5 by direct integration. Alternatively, one can use purely geometrical arguments based on the background on-shell geometry. The field equation is

$$n' = -\sqrt{H^2 + \frac{n^2}{\ell^2}}, \quad (7.2.10)$$

so the normalization requirement is

$$2\mu F^2 \int_0^1 \frac{n^2 \ell\, \mathrm{d}n}{\sqrt{H^2\ell^2 + n^2}} = 1. \quad (7.2.11)$$

This does not depend on a detailed knowledge of the form of $n$, except through $n'$. Recall that $\mu$ is the ratio $\kappa_4^2/\kappa_5^2$ of the four- and five-dimensional gravitational couplings, and $\ell$ is the AdS curvature scale, which in the case of vanishing four-dimensional cosmological constant equals $\mu^{-1}$ (cf. (5.1.12)). One now makes a trigonometric substitution to evaluate

the integral. The result is

$$\mu\ell F^2\left(\sqrt{1+H^2\ell^2}-H^2\ell^2\sinh^{-1}H\ell\right)=1. \tag{7.2.12}$$

If $\ell=\mu^{-1}$, then this agrees with Gorbunov et al. (2001); Langlois et al. (2000). In this form, the properties of $F$ are most transparent. There is an important differential equation satisfied by the normalization. One multiplies (7.2.11) by $H^2$ and differentiates logarithmically, leaving

$$\frac{\mathrm{d}\ln HF}{\mathrm{d}\ln H}+\mu F^2\frac{\mathrm{d}}{\mathrm{d}\ln H}\int_0^1\frac{n^2\ell\,\mathrm{d}n}{\sqrt{H^2\ell^2+n^2}}=1. \tag{7.2.13}$$

It is now easy to differentiate under the integral sign, and integrate the resulting expression which gives

$$\frac{\mathrm{d}}{\mathrm{d}\ln H}\int_0^1\frac{n^2\ell\,\mathrm{d}n}{\sqrt{H^2\ell^2+n^2}}=\frac{1}{\mu F^2}-\frac{\ell}{\sqrt{H^2\ell^2+1}}. \tag{7.2.14}$$

(To obtain this one needs to use the equation for $F^2$ in order to substitute for the result of integrating (7.2.13).) Substituting this into (7.2.13) gives the result

$$\mathrm{d}\ln HF=\frac{\ell\mu F^2}{(H^2\ell^2+1)^{1/2}}\,\mathrm{d}\ln H. \tag{7.2.15}$$

This relation was first observed by Huey and Lidsey (2001). It is an unusual and unexpected property of the normalization which is not enforced from first principles by any symmetry or requirement of the formalism of quantum field theory. Instead, it arises purely spontaneously, and, although (7.2.11) implies that it has a geometrical origin, no convincing geometrical interpretation has ever been offered. The important of Eq. (7.2.15) lies in the fact that it will guarantee that the consistency relation (which relates the spectral slope $n_T$ of the tensor power spectrum produced during inflation to the ratio of the amplitudes of the scalar and tensor power spectra) familiar from four-dimensional inflation also holds in the brane world. This is potentially disastrous. The consistency relation plays a pivotal role in the perturbative potential reconstruction programme (Lidsey et al., 1997) which is one of the direct aims of cosmic microwave background experiments. Its presence burdens the programme with an unwanted maze of degeneracies which could limit the accuracy and confidence with which future CMB data can be interpreted Liddle and Taylor (2002).

## 7.3. The consistency relation

We now briefly describe how the consistency relation (7.1.1) arises. Before proceeding to formal expressions, it is useful to indicate why one should expect a consistency relation

on general grounds. This is sometimes explained by saying that in the slow-roll formalism one has fewer parameters than observables, so some kind of relationship is inevitable. Although this is true, the relationship has to be developed order-by-order in perturbation theory and it is not clear why the final result is what it is. We cannot give a first principles derivation, but the following analysis is suggestive.

The amplitude of the scalar power spectrum must be related to the density perturbation $\delta$. On purely geometrical grounds, this is $A_S^2 \sim \delta^2 \sim (H/\dot{\phi})^2 \delta\phi^2$. On the other hand, since the graviton is an effective massless free field its spectrum should be $A_T^2 \sim \delta\phi^2$. The ratio $A_S^2/A_T^2$ therefore involves only geometrical quantities. Writing $H' = \mathrm{d}H/\mathrm{d}\phi$, the tensor spectral index is $n_T \sim \mathrm{d}\ln \delta\phi^2/\mathrm{d}\ln k \sim \dot{\phi}H'/H^2$, again only using geometrical considerations and enough QFT to compute the spectrum of free, massless field in de Sitter space.

So far we have not used the Einstein field equations. These are seemingly necessary to relate $A_S^2/A_T^2$ and $n_T$, since one must know how to express $\dot{\phi}$ in terms of $H$. Knowledge of this relationship is equivalent to the Friedmann equation. Although the Einstein field equations are sufficient, they are in fact stronger than necessary, because any theory of gravity which gives an action for the Friedmann-like conformally flat metric $\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{e}^{3\lambda(t)}\mathrm{d}\mathbf{x}^2$ which is proportional to the kinetic energy $\dot{\lambda}^2$ necessarily implies the Hamilton–Jacobi relation $\dot{\phi} \propto H'$ (Lidsey et al., 1997). The exact constants of proportionality are fixed by the details of the theory. In the case of Einstein gravity, one obtains

$$H' = -\frac{\kappa_4^2}{2}\dot{\phi}. \tag{7.3.1}$$

- Four dimensions. Now consider four-dimensional inflation driven by a scalar field. The matter and gravitational power spectra satisfy (4.6.85) and (4.6.89) respectively, and – as noted above – their ratio is a purely geometrical quantity,

$$\frac{A_T^2}{A_S^2} = \frac{\kappa_4^2}{2}\frac{\dot{\phi}^2}{H^2}. \tag{7.3.2}$$

The value of this quantity is set by the Friedmann equation and the classical field equation for $\phi$, so one can consider it to be a function of the type of matter under discussion and the theory of gravity being employed. In particular, if one assumes that the scalar field $\phi$ is the only constituent of the universe, then

$$\frac{A_T^2}{A_S^2} = \frac{2}{\kappa_4^2}\left(\frac{H'}{H}\right)^2 = \varepsilon, \tag{7.3.3}$$

To evaluate the tensor index, a fudge is necessary. In the idealized perfect de Sitter epoch we are considering here, the amplitude (4.6.89) is constant, since $H$ is invariant over the lifetime of the universe. Therefore the amplitude is constant on all scales, and the spectral index is zero. On the other hand, if $H$ is changing only very slowly, then a de Sitter state should be a good approximation over a small interval. The power leaving the horizon during such an interval must be close to (4.6.89) (because the corrections must involve derivatives $\dot{H}$ of the Hubble parameter, which is small), and a weak spectral index will be generated because of the variation in power as successive scales leave the horizon. Note that there is no need to retro-fit such an approximation to the exact results (4.6.85) and (4.6.89), which are calculated taking account of the expansion of $H$. In this chapter, all calculations assume a fixed Hubble rate and then make the approximation of slowly varying $H$. This is not entirely desirable, but the difficulty arises because analogous calculations to the quantum mechanical machinery leading to (4.6.89) (in particular) cannot yet be carried out.

Letting $H$ vary slowly, the first order tensor index is

$$n_T = -\frac{4}{\kappa_4^2}\left(\frac{H'}{H}\right)^2 = -2\varepsilon. \tag{7.3.4}$$

Substituting this expression into (7.3.3) immediately reproduces the consistency equation, (7.1.1).[2]

- The braneworld. In the braneworld, since the entire effect of the large extra dimensions appears as a renormalization $F$ of the tensor amplitude, the ratio $A_T^2/A_S^2$ is still independent of $A_\phi^2$.

In order to give an explicit expression, one needs to know the relationship between $H$ and $\phi$. As before, this depends only on the evolution of $\phi$ and the theory of gravity under discussion. Taking into account the relevant modifications, one finds

$$\frac{A_T^2}{A_S^2} = \frac{2F^2}{\kappa_4^2}\frac{\ell^2}{H^2+\ell^2}\left(\frac{H'}{H}\right)^2. \tag{7.3.5}$$

[2]There are directions in which this result can be generalized. See, for example, Bartolo, Matarrese, and Riotto (2001).

The tensor spectral index $n_T = \mathrm{dln}\, A_T^2/\mathrm{dln}\, k$ no longer depends merely on the functional form of $A_\phi^2$, which is unchanged from the four-dimensional case, but instead receives non-trivial corrections from the renormalization $F^2$. In particular, $\mathrm{d}\ln A_T^2 = 2\mathrm{d}\ln HF$. At first sight, this would appear to break any hope of retaining the consistency relation. However, (7.2.15) combined with (7.3.5) ensures that the consistency relation survives.

Although we have presented this result only in the case of a pure anti-de Sitter bulk with $\mathbf{Z}_2$ symmetry, the appearance of the consistency relation holds rather more generally. In particular, Huey and Lidsey (2001) have argued that it persists if one allows different anti-de Sitter curvatures $\ell_<$, $\ell_+$ on the $y < 0$ and $y > 0$ branches.

## 7.4. Fluctuations in a perturbed four-dimensional de Sitter space

**7.4.1. Introduction.** In this section, we aim to calculate the power spectrum of a scalar field propagating over a background de Sitter cosmology with some first order perturbation, restricting the calculation to purely four dimensions at this stage. Let $\mathcal{M}$ be four-dimensional de Sitter space with fixed, time-independent Hubble parameter $H_0$. Consider a small perturbation $\Delta H$ of the Hubble rate, with arbitrary time dependence, where $\Delta H$ is supposed to be sufficiently small that terms quadratic or higher in $\Delta H$ can be ignored. We wish to calculate the power spectrum of a free, massless scalar field propagating on this fixed geometry.

The study of inflationary fluctuations is quite mature and a number of effects are routinely included in calculations (Lidsey et al., 1997; Stewart and Lyth, 1993). An important example is the coupling of metric perturbations to $\phi$, since observations are approaching the precision at which one should include next-order effects. It is well known that in pure de Sitter space there is, in fact, no coupling: any metric fluctuations are pure gauge. To obtain coupling between scalar field fluctuations and bulk metric perturbations, one must have some measure of tilt away from de Sitter space. This tilt is precisely what is measured by the slow-roll parameters $\varepsilon$ and $\eta$, and their higher order relatives. If $\varepsilon = 0$, then one is in exact de Sitter space, and there are no metric fluctuations; if $\varepsilon \neq 0$, then one should take account of the coupling. In the present case, bearing in mind the order to which we

carry perturbation theory, there is no coupling between the scalar field perturbation and gravitational fluctuations.

**7.4.2. Scalar and tensor power spectra.** The two-point function $G(x_1, x_2) = \mathrm{i}\langle\phi(x_1)\phi(x_2)\rangle$ is given by (7.2.1), so its Fourier transform,

$$G(x_1, x_2) = (2\pi)^{-3} \int \mathrm{d}^3k\, G(t_1, t_2; k) \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{x}_1 - \mathbf{x}_2)} \tag{7.4.1}$$

should satisfy

$$\left(\frac{\partial^2}{\partial {t_1}^2} + 3H\frac{\partial}{\partial t_1} + \frac{k^2}{R^2}\right) G(t_1, t_2; k) = \frac{\delta_{\mathrm{D}}(t_1 - t_2)}{R^3}. \tag{7.4.2}$$

This can be solved to first order in $\Delta H$. Thus, one is trying to build a Green's function $G = G_0 + G_1$, where $G_0$ is the background Green's function and $G_1$ is a perturbation. It is convenient to change variable to the conformal time, $\mathrm{d}\tau = R_0^{-1}\,\mathrm{d}t$, and set $G_n = u_n/R_0$ for $n = 0, 1$. One separates (7.4.2) into zero- and first-order parts. The zero-order part is Mukhanov's equation for the background,

$$\left(\frac{\partial^2}{\partial {\tau_1}^2} + k^2 - 2(R_0H_0)^2\right) u_0 = \frac{\delta_{\mathrm{D}}(\tau_1 - \tau_2)}{R_0} \tag{7.4.3}$$

and the first order part is a sourced Mukhanov equation,

$$\left(\frac{\partial^2}{\partial {\tau_1}^2} + k^2 - 2(R_0H_0)^2\right) u_1 = 2k^2\frac{\Delta R}{R_0}u_0 - 3\Delta H(R_0u_0' - u_0H_0R_0^2) - 3\frac{\Delta R}{R_0}\frac{\delta_{\mathrm{D}}(\tau_1 - \tau_2)}{R_0}, \tag{7.4.4}$$

at first order, where $\delta_{\mathrm{D}}$ is the Dirac delta-function. The unperturbed Green's function $G_0$ can be thought of as the Feynman propagator for the free theory defined by the de Sitter background, and $G_1$ as the first term in a Feynman series for interactions introduced by the departure of the background from exact de Sitter. Mukhanov's equation has a well-known solution (Birrell and Davies, 1982),

$$G_0 = \frac{\mathrm{i}\pi}{4k}\frac{1}{R_0(\tau_1)R_0(\tau_2)}L^{(1)}(-k\tau_1)L^{(2)}(-k\tau_2), \tag{7.4.5}$$

if $\tau_2 > \tau_1$, or the same expression with $\tau_1$ and $\tau_2$ exchanged if not. The functions $L^{(1,2)}$ are defined in (4.6.72), and the result has been written in terms of $G$ in order to exhibit the symmetry between $\tau_1$ and $\tau_2$. The boundary conditions as $|k\tau| \to \infty$ are fixed by the Bunch–Davies vacuum.

The remaining obstacle is the first-order sourced Mukhanov equation (7.4.4). To solve this, one can choose either to employ some generally applicable technology, such as standard Sturm–Liouville theory (Morse and Feshbach, 1953), or seek direct solutions. For technical

reasons we prefer the Sturm–Liouville approach, although a direct integral solution can also be given.[3] We define eigenfunctions for the background equation, of weight $-m^2$,

$$\left[\frac{\partial^2}{\partial\tau^2} + \left(k^2 - \frac{2}{\tau^2}\right)\right]\Omega_m = -m^2\Omega_m. \tag{7.4.13}$$

One must impose sufficient boundary conditions to make the $\Omega_m$ behave well, after which the background field equation and the $\Omega_m$ form a self-adjoint set. The boundary condition on the $\Omega_m$ at $\tau = -\infty$ is expected to be immaterial, provided the $\Omega_m$ decay sufficiently fast there. At $\tau = 0$, we demand that the $\Omega_m$ be regular. This selects the Bessel function,

$$\Omega_m(k,\tau) = \sqrt{-K\tau}J_{3/2}(-K\tau), \tag{7.4.14}$$

---

[3] One is trying to solve the equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}\tau^2} + k^2 - \frac{2}{\tau^2}\right)u(\tau) = f(\tau) \tag{7.4.6}$$

for some source function $f(\tau)$. One can integrate to find

$$\begin{aligned} u(\tau) = -\frac{1}{\tau k^{3/2}}\Bigg[&\alpha S^{(1)}(k\tau) + \beta S^{(2)}(k\tau) \\ &+ \int_{-\infty}^{\tau}\frac{f(x)\,\mathrm{d}x}{xk^{3/2}}\left(S^{(1)}(k\tau)S^{(2)}(kx) - S^{(1)}(kx)S^{(2)}(k\tau)\right)\Bigg], \end{aligned} \tag{7.4.7}$$

where the functions $S^{(1)}(z)$ and $S^{(2)}(z)$ are related to the Hankel functions of order 3/2,

$$S^{(1)}(z) = \cos z + z\sin z, \quad S^{(2)}(z) = \sin z - z\cos z. \tag{7.4.8}$$

For calculating the perturbed power spectrum, the appropriate source function is

$$f = 2k^2\frac{\Delta R}{R_0}u_0 - 3\Delta H(R_0u_0' - R_0^2H_0u_0). \tag{7.4.9}$$

By proceeding with the standard argument by which one calculates the power spectrum, one arrives at the expression

$$A_\phi^2 = \frac{H_0^2}{4\pi^2}\left(1 - 3\frac{\Delta R}{R_0} - \frac{\mathrm{i}}{k^2}\sqrt{\frac{\pi}{2}}\int_{-\infty}^{\tau}\frac{\mathrm{d}\sigma}{\sigma}\,\bar{S}(\tau,\sigma)C(\sigma)\right), \tag{7.4.10}$$

where

$$\bar{S} = (1 - k^2\tau\sigma)\sin k\sigma - k(\tau+\sigma)\cos k\sigma \tag{7.4.11}$$

and

$$\begin{aligned} C =&2k^2\frac{\Delta R}{R_0}(-k\sigma)^{1/2}H_{3/2}^{(1)}(-k\sigma) \\ &+ 3\Delta H\left(\frac{3}{2H_0\sigma^2}H_{3/2}^{(1)} - \frac{k}{H_0\sigma}H_{3/2}^{(1)\prime}\right), \end{aligned} \tag{7.4.12}$$

where the argument of each Hankel function is $-k\sigma$. The relationship between this solution and the Liouville transform solution presented in the main text is the same as the relationship between (for example) Laplace transform solutions and solutions provided by the "operational method" — see Jeffreys and Jeffreys (1946).

where $K^2 = k^2 + m^2$ as before. The $\Omega_m$ obey an orthonormality relation on $\tau \in (-\infty, 0]$

$$\int_{-\infty}^{0} \mathrm{d}\tau\ \Omega_m(\tau)\Omega_n(\tau) = \delta_{\mathrm{D}}(m-n) \tag{7.4.15}$$

or, equivalently, $\int_{-\infty}^{\infty} \mathrm{d}m\ \Omega_m(\tau)\Omega_m(\sigma) = \delta(\tau-\sigma)$. The background equation on $y \in [0, y_{\mathrm{h}}]$ is singular in the Sturm–Liouville sense, so there is a continuum of eigenvalues and not simply a discrete set. Using the $\Omega_m$ and the completeness relation, the part of $u_1$ given by the driving term in (7.4.4) can be solved

$$u_1^{\mathrm{drive}}(\tau_1, \tau_2) = -\int_{-\infty}^{0} \mathrm{d}\sigma\ \mathsf{K}(\tau_1, \sigma)U(\sigma, \tau_2) \tag{7.4.16}$$

where the transition matrix kernel $\mathsf{K}(\tau_1, \tau_2)$ is

$$\mathsf{K}(\tau_1, \tau_2) = \int_{-\infty}^{\infty} \frac{\mathrm{d}m}{m^2}\ \Omega_m(\tau_1)\Omega_m(\tau_2), \tag{7.4.17}$$

and $U(\sigma, \tau_2)$ satisfies

$$U(\sigma, \tau_2) = -2k^2\frac{\Delta R}{R_0}(\sigma)u_0(\sigma, \tau_2) - 3\Delta H(\sigma)\left[R_0(\sigma)u_0'(\sigma, \tau_2) - R_0^2(\sigma)H_0 u_0(\sigma, \tau_2)\right]. \tag{7.4.18}$$

From (7.4.14) it is clear that (7.4.16) is no more than solution via a Fourier–Bessel transform (Jeffreys and Jeffreys, 1946; Riesz and Sz.-Nagy, 1955). There is also an impulsive contribution to $u_1$ arising from the $\delta$-function. Assuming one makes the same choice of boundary conditions, which is necessary in order to prevent disturbance of the asymptotic quantum vacuum, this must be just proportional to $u_0$, giving

$$u_1^{\mathrm{impulse}}(\tau_1, \tau_2) = -3\frac{\Delta R}{R_0}(\tau_2)u_0(\tau_1, \tau_2) \tag{7.4.19}$$

One can now assemble $G_0$ and $G_1$ to construct the full two-point function, restoring the necessary factors of $\exp[\mathrm{i}\mathbf{k}\cdot(\mathbf{x}_1 - \mathbf{x}_2)]$ and integrations over $\mathbf{k}$. We find

$$G(x_1, x_2) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, W(\tau_1, \tau_2; k)\, \exp[\mathrm{i}\mathbf{k}\cdot(\mathbf{x}_1 - \mathbf{x}_2)] \tag{7.4.20}$$

where $W$ satisfies

$$\begin{aligned} W(\tau_1, \tau_2; k) = {} & \frac{\mathrm{i}\pi}{4k}\frac{1}{R(\tau_1)R(\tau_2)}\left[1 - 3\frac{\Delta R}{R_0}(\tau_2)\right]L^{(1)}(-k\tau_2)L^{(2)}(-k\tau_1) \\ & - \frac{1}{R_0(\tau_1)}\int_{-\infty}^{0} \mathrm{d}\sigma\ \mathsf{K}(\tau_1, \sigma)U(\sigma, \tau_2), \end{aligned} \tag{7.4.21}$$

if $\tau_1 > \tau_2$, and the same expression with $\tau_1$ and $\tau_2$ exchanged in the first term otherwise. To find the power spectrum one lets $\mathbf{x}_1$ and $\tau_1$ approach $\mathbf{x}_2$ and $\tau_2$, respectively, and takes

a logarithmic derivative with respect to $k$. The result is

$$A_\phi^2 = -\mathrm{i}\frac{k^3}{2\pi^2}W(\tau,\tau;k). \tag{7.4.22}$$

We take the limit $k/RH \to 0$ to give the asymptotic behaviour on large scales (Lidsey et al., 1997). One must divide the $\sigma$ integral into two regions, where $\sigma < \tau_2$ and $\sigma > \tau_2$ respectively. We find

$$A_\phi^2 \overset{k/RH\to 0}{\longrightarrow} \frac{H_0^2}{4\pi^2}\left(1 - 3\frac{\Delta R}{R_0} - \sqrt{\frac{\pi}{2}}k\tau Q(\tau)\right), \tag{7.4.23}$$

where $Q(\tau)$ is defined by a convolution with the transition matrix kernel $\mathsf{K}$,

$$Q(\tau) = \lim_{k/RH\to 0}\int_{-\infty}^{0} \mathrm{d}\sigma\, \mathsf{K}(\sigma,\tau)\mathcal{L}(\sigma,\tau)\left[2k^2\frac{\Delta R}{R_0} + 3\frac{\Delta H}{H_0}\left(\frac{3}{2\sigma^2} - \frac{3k}{\sigma}\mathcal{H}(\sigma,\tau)\right)\right]. \tag{7.4.24}$$

The auxiliary quantities $\mathcal{L}$ and $\mathcal{H}$ satisfy

$$\mathcal{L}(\sigma,\tau) = \begin{cases} \mathrm{i}L^{(1)}(-k\sigma) & \tau > \sigma \\ -\mathrm{i}L^{(2)}(-k\sigma) & \tau < \sigma \end{cases} \tag{7.4.25}$$

and

$$\mathcal{H}(\sigma,\tau) = \begin{cases} H_{3/2}^{\prime(1)}(-k\sigma)/H_{3/2}^{(1)}(-k\sigma) & \tau > \sigma \\ H_{3/2}^{\prime(2)}(-k\sigma)/H_{3/2}^{(2)}(-k\sigma) & \tau < \sigma \end{cases} \tag{7.4.26}$$

The function $Q$ will appear frequently. It is non-local and represents the effect of time-dependence in our perturbation.

When comparing (7.4.23) with similar models in the literature, one should bear in mind that this expression constitutes a global solution, which does not involve a power series expansion around any particular point on the potential. To recover a standard result, one can expand in Taylor series around a value $H_0 = H(\phi_{\mathrm{cl},0})$. By doing so, one can recover only the lowest order (de Sitter space) result from (7.4.23), since only first order departures from de Sitter space are included. For comparison, the leading order term in the standard analysis is $\sim \varepsilon \approx (H'/H)^2$ with would be at second order in our calculation scheme. Therefore, although the motivation is in the same in each case, the present result is not directly related to the Stewart–Lyth next-order calculation (Stewart and Lyth, 1993). In particular, the Stewart–Lyth procedure involves a power series expansion of the potential, which is locally matched onto an exact solution, and explicitly quantizes the matter fluctuations. This does not happen in our approach: there is no coupling to matter fluctuations. This compromise does not represent the calculation

one would ideally wish to carry out, but it does represent the only meaningful generalization which can practically be calculated in the braneworld. The main virtue of our approach is the non-perturbative treatment of the time dependence. Of course, the result is still perturbative in the amplitude.

Using the gauge-invariant scalar variable $\zeta$, defined in (4.6.13), gives a final expression for the power spectrum of scalar curvature perturbations. When evaluated on the horizon scale, $-k\tau = 1$, we obtain an expression for the asymptotic amplitude, in terms of the values quantities had at horizon crossing:

$$A_S^2 = \frac{1}{25\pi^2}\frac{H_0^4}{\dot{\phi}_{\rm cl}^2}\left(1 + 2\frac{\Delta H}{H_0} - 3\frac{\Delta R}{R_0} + \sqrt{\frac{\pi}{2}}Q(-k^{-1})\right), \tag{7.4.27}$$

where the extra term involving $\Delta H/H_0$ arises from the use of the full perturbed Hubble rate, rather than $H_0$, in (4.6.13).

The case of gravitational waves is similar, and in fact the reasoning applied in Section 4.6.2 is still relevant: the action for each polarization of the graviton is the same as the free, massless scalar field action except that the relative normalization differs by a factor $(4\kappa_4^2)^{-1}$. There are two polarization modes, so (4.6.89) governs the tensor power spectrum,

$$A_T^2 = \frac{\kappa_4^2}{50}\frac{H_0^2}{\pi^2}\left(1 - 3\frac{\Delta R}{R_0} + \sqrt{\frac{\pi}{2}}Q(-k^{-1})\right). \tag{7.4.28}$$

The ratio of tensor to scalar amplitudes satisfies

$$\frac{A_T^2}{A_S^2} = \frac{\kappa_4^2}{2}\frac{\dot{\phi}_{\rm cl}^2}{H_0^4}\left(1 - 2\frac{\Delta H}{H_0}\right) \simeq \varepsilon\left(1 - 2\frac{\Delta H}{H_0}\right), \tag{7.4.29}$$

and the tensor spectral index is

$$n_T = -2\varepsilon\left(1 - \frac{\Delta H}{H_0}\right) - 3\frac{\Delta H}{H_0} + \sqrt{\frac{\pi}{2}}\frac{\mathrm{d}}{\mathrm{d}\ln k}Q(-k^{-1}). \tag{7.4.30}$$

One might enquire whether or not this set of amplitudes and the tensor spectral index do not already break the consistency relation. In fact this is possibly true (the consistency relation being a result known only to next-order in the slow-roll expansion (Lidsey et al., 1997) and not a non-perturbative theorem), although the complexity of the function $Q$ rather precludes the possibility of a direct verification. However, this is not the main point of the present formalism. In the next section we will argue that although we cannot specify whether the consistency relation holds in the brane world, in its usual form *or* in broken

form, this is not necessary, because the dependence of these quantities on the brane tension $\lambda$ means that they could be equal in the brane world only in fine-tuned scenario.

### 7.5. Fluctuations in a perturbed de Sitter braneworld

We now repeat the same calculation in the braneworld. Consider a de Sitter brane with Hubble parameter $H_0$ immersed in anti-de Sitter space and allow small fluctuations $\Delta\rho$ in the matter density. These fluctuations are taken to vary with the brane tension $\lambda$ in such a way as to keep $\Delta H$ the same, regardless of the value of $\lambda$. We define this to be our notion of the 'same' perturbation in the brane world and in four dimensions. As one sends $\lambda \to \infty$ one should recover the four-dimensional result with this choice of $\Delta H$, which is an expectation we will explicitly verify later.

Firstly, consider some free, massless scalar field $\phi$ propagating over the brane $\Sigma$. This theory is just the same as one would find in four dimensions, provided scalar fluctuations coming from the bulk are ignored. (As discussed above, there are circumstances under which this may not be a good approximation.) The resulting fluctuation equation will coincide with (7.4.2) and the power spectrum one derives will equal (7.4.27), provided that $R$ is taken to satisfy the expansion law for the on-brane cosmological scale factor. This is equivalent to supposing that the Mukhanov equation remains a valid description of the perturbation on the brane (Ramírez and Liddle, 2003). Although this is a sensible conjecture, there does not yet appear to be any convincing direct derivation to support it.

The case of gravitational waves is not the same. In a general geometry, the graviton wave operator $\Box$ couples the $t$ and $y$ dependence of the graviton $\mathbf{k}$-modes, so that an explicit solution is extremely difficult. One can always work on the brane universe in black hole coordinates (Bowcock et al., 2000; Mukohyama et al., 2000), where the metric is explicitly stationary, and one recovers ordinary differential equations. Unfortunately, the boundary conditions are non-trivial to apply, because the brane appears as an arbitrarily curved figure. In this section we make progress by a different route.

We begin by rewriting the general formula for $n(t, y)$ in terms of $H'$, assuming no dark radiation, where

$$\dot{H} = \dot{\phi}H' = -\frac{2\ell}{\kappa_4^2}\frac{H'^2}{\sqrt{H^2 + \ell^{-2}}}. \tag{7.5.1}$$

Since $\dot{H} \propto H'^2$, if we perturb around the de Sitter solution then the term $H' = \Delta H'$ is small and we may neglect its square. Hence, for a perturbed de Sitter brane, we still

retain $n = a/a_{\rm b}$. We emphasize that this is only true for perturbations around de Sitter space supported by a single scalar field where the background $H$ satisfies $H' = 0$. When calculating spectral indices we will again endow $H$ with very weak time dependence, but for the purposes of the calculation presented in this section the background $H$ is to be regarded as fixed, in analogy with the four dimensional calculation of $A_\phi^2$.

Let us write the metric functions $a$ and $n$ assuming no dark radiation, but not necessarily with $H$ constant, as long as $\dot{H}$ can be ignored,

$$n(y) = \frac{H\ell}{\sqrt{2}} \left[\cosh 2\ell^{-1}(y_{\rm h} - y) - 1\right]^{1/2} . \tag{7.5.2}$$

The other function $a$ satisfies $a(t, y) = a_{\rm b}(t)n(y)$ where $a_{\rm b}(t)$ is the scale factor on the brane. Writing $n$ in terms of background plus perturbed quantities $H \mapsto H_0 + \Delta H$, the function $n$ acquires an explicit time dependence, and $n(t, y)$ becomes

$$n(t, y) \mapsto \left(1 + \frac{\Delta H}{H_0}\right) \frac{H_0\ell}{\sqrt{2}} \left[\cosh 2\ell^{-1}(y_{\rm h} + \Delta y_{\rm h} - y) - 1\right]^{1/2} , \tag{7.5.3}$$

since the horizon location $y_{\rm h}$ depends on time via $H$. The perturbed $n$, Eq. (7.5.3), has the same values at each end-point as the unperturbed $n_0$, so it satisfies $n(t, 0) = 1$ and $n(t, y_{\rm h}) = 0$. This may appear surprising, because one would typically expect a perturbation to disturb these values. The condition $n(t, y = 0) = 1$ arises because of the gauge condition which fixes $t$, and as a result the perturbation in the cosh term exactly cancels the perturbation in the pre-factor at $y = 0$. The second occurs because $y = y_{\rm h}$ is a minimum of $n_0$, so it is not displaced to first order. If more terms were retained in the perturbation expansion, or any dark radiation were to be present, then $n(t, y_{\rm h})$ would change.

Any perturbation of $H$ in a four-dimensional cosmology is necessarily sourced by a corresponding change in the matter density $\rho$, in virtue of the four-dimensional Friedmann equation. This simplicity does not carry over to the brane world. Instead, the possible existence of a dark radiation component allows a one-parameter family of choices, all of which can source any given $\Delta H$. This allows us to identify two distinct perturbation modes, which we designate Type I and Type II: the first corresponding to a perturbation of the density $\rho$ which leaves the dark radiation $\mathcal{C}$ intact, and the second corresponding to the opposite arrangement. A general perturbation will be some admixture of the two. Recall that in our geometry, the dark radiation component is initially absent. To proceed it is necessary to decide how $\Delta H$ is to be split between $\rho$ and $\mathcal{C}$.

The presence of a dark radiation component $\mathcal{C}$ will presumably change the physics, since it involves the introduction of an extra tunable parameter in the description. For this reason we would like it to be absent, because in the four-dimensional model the perturbation came entirely from the matter sector. In order to achieve a proper comparison with the braneworld result, the perturbation here should also arise entirely from density perturbations and not the introduction of dark radiation. Dark radiation is equivalent to Weyl curvature in the bulk spacetime (Binetruy et al., 2000a). One can measure such curvature by any suitable invariant formed from the Weyl tensor $C_{abcd}$, which is the part of the Riemann tensor not determined by the Ricci curvature. It can be described as the "free" part of the gravitational field. For example, one can choose the square of the Weyl tensor, $\Psi = C^{abcd}C_{abcd}$. By allowing $H$ and $y_{\rm h}$ to vary with time, while keeping the general form of the solution (7.5.2), one finds a Weyl invariant of the form

$$\Psi = \frac{2}{H^8}\left[\ell^{-4}\,\mathrm{cosech}\,\mu(y_{\rm h}-y)\right]^8\Big((4\dot{H}^2-2H\ddot{H})\sinh^2\ell^{-1}(y_{\rm h}-y)+ H(2\ell^{-1}\dot{H}\dot{y_{\rm h}}+\ell^{-1}H^2\dot{y_{\rm h}}-\ell^{-1}H\ddot{y_{\rm h}})\sinh 2\ell^{-1}(y_{\rm h}-y)+ \ell^{-2}H^2\left[3+\dot{y_{\rm h}}^2\cosh 2\ell^{-1}(y_{\rm h}-y)\right]\Big)^2. \tag{7.5.4}$$

$\Psi$ has a leading contributions proportional to $\dot{y_{\rm h}}^2$. Since $\Psi$ is quadratic in $C_{abcd}$, this means that $C_{abcd}$ itself is quadratic in $\delta H$ and therefore zero at first order. Alternatively, one can see that since $\mathcal{C}$ is zero in the unperturbed geometry, it can enter only at $\delta\dot{H}$ in the perturbed cosmology. Since we are ignoring time variation in $\delta\dot{H}$ as a second order effect, no Weyl curvature, or dark radiation, is induced to leading order by the perturbation. Therefore, although there is no reason of principle why Type II perturbations should not be present, our future considerations will be restricted to cases where they are not. It should be noticed that there appears to be no known analytic solution, either perturbative or non-perturbative, for the form of the gravitational wavefunction in the presence of dark radiation.

**7.5.1. The tensor zero mode.** We now solve for the graviton zero mode. The method of analysis applied in the unperturbed case, based on a standard decomposition of the path integral action into harmonics of the transverse dimension no longer makes sense here, because the metric functions, such as (7.5.3), no longer separate. Our analysis

is based on a specific Ansatz: we suppose that the graviton zero mode remains distinct, and carries no dependence on the transverse dimension.

This discussion of toplogical stability of the zero mode under metric deformations of the compactification manifold which was set out in Chapter 5 does not, unfortunately, supply an *a priori* guarantee that this is sensible. On the one hand, whenever the brane compactification can be given a bundle structure, as a fibration of our universe over the compact space, the standard results about topological stability do suggest that the zero mode should be stable. (This can equivalently be expressed by saying that the five-dimensional wave operator separates.) On the other hand, this perturbation is carrying the compactification away from the under-control bundle compactification scenario and into a more general class of metrics. This is ultimately responsible for the failure of the Kaluza–Klein expansion, and by the same stroke invalidates the assignment of zero modes to cohomology classes. To understand why this supposition works requires more analysis. Consider the classical field equation for the graviton, $\Box\Psi = 0$,

$$\left(-\frac{1}{n^2}\frac{\partial^2}{\partial t^2} - \frac{\omega}{n^2}\frac{\partial}{\partial t} + \frac{\triangle}{a^2} + \frac{\partial^2}{\partial y^2} + \sigma\frac{\partial}{\partial y}\right)\Psi = 0, \tag{7.5.5}$$

where the coefficient functions $\omega$ and $\sigma$ are given by

$$\omega = 3\frac{\dot{a}}{a} - \frac{\dot{n}}{n} \quad \text{and} \quad \sigma = 3\frac{a'}{a} + \frac{n'}{n}. \tag{7.5.6}$$

This is to be expanded to first order in the perturbations $\Delta a$ and $\Delta n$. The explicit solution to the background equation was discussed in Chapter 5, and has previously appeared in the literature (Gorbunov et al., 2001; Langlois et al., 2000). At first order one obtains, writing $\Psi = \Psi_0 + \Psi_1$ for the expansion of the field,

$$\Psi_1'' + 4\frac{n_0'}{n_0}\Psi_1' - \frac{1}{n_0^2}\ddot{\Psi}_1 - \frac{3H_0}{n_0^2}\dot{\Psi}_1 - \frac{k^2}{a_0^2}\Psi_1 = \frac{\Delta\omega}{n_0^2}\dot{\Psi}_0 - \frac{2}{n_0^2}\frac{\Delta n}{n_0}\left(\ddot{\Psi}_0 + 3H_0\dot{\Psi}_0\right) - \frac{2k^2}{a_0^2}\frac{\Delta a}{a_0}\Psi_0 - \Delta\sigma\Psi_0'. \tag{7.5.7}$$

One can show that $\Delta\omega = 3\Delta H$. Restricting attention to the perturbation of the zero mode and making use of the background field equation, this becomes

$$\Psi_1'' + 4\frac{n_0'}{n_0}\Psi_1' - \frac{1}{n_0^2}\ddot{\Psi}_1 - \frac{3H_0}{n_0^2}\dot{\Psi}_1 - \frac{k^2}{a_0^2}\Psi_1 = \frac{3\Delta H}{n_0^2}\dot{\Psi}_0 + \frac{2k^2}{R_0^2 n_0^3}\Psi_0\left(\Delta n - \frac{\Delta a}{R_0}\right). \tag{7.5.8}$$

The right hand side appears to be a complicated function of $t$ and $y$. This is true in general, but the special relationship $a = Rn$ reduces the term in brackets to $n_0\Delta R/R_0$; this is a

consequence of the assumption that $\dot{H}$ vanishes. In virtue of this simplification, we can separate the field equation into a $y$-derivative piece

$$\Psi_1'' + 4\frac{n_0'}{n_0}\Psi_1' = 0 \tag{7.5.9}$$

and a $t$-derivative piece

$$\ddot{\Psi}_1 + \frac{3H_0}{n_0^2}\dot{\Psi}_1 + \frac{k^2}{R_0^2}\Psi_1 = \frac{2k^2}{R_0^2}\Psi_0\frac{\Delta R}{R_0} - 3\Delta H\dot{\Psi}_0. \tag{7.5.10}$$

More generally there should be a separation constant relating these two equations, which we have set to zero in the present case to pick out the zero mode. This 'unwrapping' or local trivialization of the compactification happens only for the zero mode, and not for any higher mass modes. Although it is not obviously topological in character, it does emphasize the priviledged status of the zero-mode.[4] As in the standard case, the $y$ equation has solutions $\Psi_1 \propto$ constant or $\Psi_1 \propto n_0^{-4}$. Since $n_0 \to 0$ at the Cauchy horizon, the correct solution is to take the $y$-dependence of $\Psi_1$ to be constant, preventing an unwanted divergence at $y = y_{\rm h}$. This choice is a necessary consequence of the boundary conditions on $\Psi$, which enforce $\Psi' = 0$ at the horizon in order to keep anisotropic stress absent.

On its own, this calculation is insufficient to obtain the two-point function for the zero-mode, which should properly be obtained from a functional integral like (7.2.1). Consider the two-point function for a polarization mode $\phi$ of the graviton, and suppose we can split $\phi$ into a zero-mode piece $\phi_0$, or collective excitation, which has no transverse dependence, and an unimportant remainder which encodes the details of heavy Kaluza–Klein modes. We assume it is permissible to ignore these heavy modes. In order for this procedure to make sense, we must suppose that the zero-mode is stable under small perturbations. The two-point function becomes

$$\langle\phi(x_1)\phi(x_2)\rangle = \int [\mathrm{d}\phi_0]\ \phi(x_1)\phi(x_2)\ \exp\left(-\frac{\mathrm{i}}{8\kappa_5^2}\int \mathrm{d}v\ \phi_0\Box\phi_0\right). \tag{7.5.11}$$

Since $\phi_0$ has no transverse dependence by assumption, the action of the braneworld Laplacian $\Box$ on $\phi_0$ is the same as the de Sitter Laplacian $\Box_{\rm dS}$, where

$$\Box_{\rm dS} = -\frac{\partial^2}{\partial t^2} - 3H\frac{\partial}{\partial t} + \frac{\triangle}{R^2}. \tag{7.5.12}$$

[4]Recall that the zero mode did not appear in the spectrum based on the boundary behaviour, or conformal scaling dimension, argument in Chapter 5.

Dropping the 0 subscript on $\phi_0$, the action for $\phi$ must be

$$S = \int \mathrm{d}x\ \phi\Box\phi = \int \mathrm{d}^3x\,\mathrm{d}t\,\mathrm{d}y\ R^3 n^2\,\phi\Box_{\mathrm{dS}}\phi. \tag{7.5.13}$$

One now integrates over $y$ to obtain an effective four-dimensional action, which, since the only $y$ dependence occurs in $n$, must be of the form

$$S = \int \mathrm{d}^3x\,\mathrm{d}t\,\mathrm{d}y\ \left(n_0^2 R^3 + 2n_0\frac{\Delta n_0}{\Delta H}R_0^3\Delta H\right)\ \phi\Box_{\mathrm{dS}}\phi. \tag{7.5.14}$$

This can be split in two, and each integral performed separately. The integral $\int \mathrm{d}y\, n_0^2$ is just the familiar normalization factor $(\mu F^2)^{-1}$. The new contribution from the perturbed piece is, explicitly,

$$Y^2 = 4\int_0^{y_{\mathrm{h}}} \mathrm{d}y\ n_0\frac{\Delta n}{\Delta H} = H_0\ell^3\left(\frac{2(1+\cosh\ell^{-1}y_{\mathrm{h}})}{\sqrt{1+H_0^2\ell^2}} + \pi + 4\arctan \mathrm{e}^{\ell^{-1}y_{\mathrm{h}}} - 2\sinh\ell^{-1}y_{\mathrm{h}}\right). \tag{7.5.15}$$

$Y^2$ has a simple geometrical interpretation. In the background geometry, the brane and the horizon are parallel. Integrating over the volume between them, with the correct AdS measure, gives the normalization $\mu F^2$. When the perturbation is introduced, the volume of the AdS slice between the brane and the horizon is changed, because of the deformation suffered by the metric function $n$. The extra normalization piece $Y^2$ takes account of this change in volume. A subtle feature is that the background normalization should be integrated between $y = 0$ and the real horizon at $y = y_{\mathrm{h}} + \Delta y_{\mathrm{h}}$, but in fact it is easy to see that this introduces no new terms, because $n_0(t, y = y_{\mathrm{h}}) = 0$. Therefore, we are entitled to carry all volume integrals only to the unperturbed horizon, at $y = y_{\mathrm{h}}$.

This understanding of the origin of $Y^2$ provides a useful physical characterization of the approximation that $\dot{H} = 0$, whose ramifications are not obvious merely from inspection of the formulae for $n$ and $a$. The physical content of this approximation is that we are including only the 'breathing mode' of the perturbation. In particular, couplings of the wave zero mode to curvature fluctuations in the bulk are neglected; a more sensitive analysis will be needed to decide if such couplings play an important role.

Combining the two integrals appearing here gives the four-dimensional effective action correct to first order,

$$\langle\phi(x_1)\phi(x_2)\rangle = \int[\mathrm{d}\phi]\ \phi(x_1)\phi(x_2)\exp\left(-\frac{\mathrm{i}}{8\kappa_5^2}\frac{1}{\mu F^2}\int_\Sigma \mathrm{d}^3x\,\mathrm{d}t\ R^3(1+\mu F^2Y^2\Delta H)\ \phi\Box_{\mathrm{dS}}\phi\right), \tag{7.5.16}$$

where $x_1$ and $x_2$ are taken to lie on the brane. This is now amenable to solution using the four-dimensional methods of the preceding section. The two-point function satisfies

$$\langle\phi(x_1)\phi(x_2)\rangle = -4\mathrm{i}\kappa_4^2 F^2 G(x_1, x_2), \tag{7.5.17}$$

where $G(x_1, x_2)$ is the Green's function for $\Box_{\mathrm{dS}}$ in the measure $R^3(1+\mu F^2 Y^2 \Delta H)$. Therefore $G(t_1, t_2; k)$ should solve (cf. (7.4.2))

$$\left(\frac{\partial^2}{\partial {t_1}^2} + 3H\frac{\partial}{\partial t_1} + \frac{k^2}{R^2}\right) G_0(t_1, t_2; k) = \left(1 - \mu F^2 Y^2 \delta H\right) \frac{\delta_{\mathrm{D}}(t_1 - t_2)}{R^3}. \tag{7.5.18}$$

It is now immediate that the tensor power spectrum satisfies

$$A_T^2 = \frac{\kappa_4^2 F^2 H_0^2}{50\pi^2}\left(1 - 3\frac{\Delta R}{R_0} - \mu H_0 F^2 Y^2 \frac{\Delta H}{H_0} + \sqrt{\frac{\pi}{2}} Q(-k^{-1})\right), \tag{7.5.19}$$

where $F$ involves the background Hubble rate $H_0$ only. One should check that this expression has the correct form in the decoupling limit $\lambda \to \infty$ where the brane tension diverges. In the present case, this is equivalent to $\ell \to 0$ (see Chapter 5), so it is easy to verify that $Y^2 \to 0$, and $A_T^2$ reverts smoothly to its four-dimensional equivalent. In showing this, it is essential that $\sinh \ell^{-1} y_{\mathrm{h}}$ and $\cosh \ell^{-1} y_{\mathrm{h}}$ diverge only like $\ell^{-1}$ as $\ell \to 0$.

**7.5.2. Braneworld consistency relation.** In five dimensions there an obstruction to any attempt to re-establish the consistency relation. This obstruction arises from the change in normalization of the graviton zero mode, and in particular its dependence on the brane tension $\lambda$. To see how this works in detail, we make the approximation that to calculate the tensor spectral index one takes $H_0$ to be a slowly rolling function of $\phi$,

$$H_0^2 + 2H_0 \Delta H_0 = \frac{\kappa_4^2}{3}\rho_\phi\left(1 + \frac{\rho_\phi}{2\lambda}\right) + \frac{\kappa_4^2}{3}\Delta\rho\left(1 + \frac{\rho_\phi}{\lambda}\right). \tag{7.5.20}$$

This is just the perturbed Friedmann equation. Therefore the perturbation $\Delta H$ satisfies

$$\Delta H = \frac{\kappa_4}{2\sqrt{3}} \frac{\Delta\rho}{\rho_\phi^{1/2}}\left(1 + \frac{\rho_\phi}{\lambda}\right)\left(1 + \frac{\rho_\phi}{2\lambda}\right)^{-1/2} \tag{7.5.21}$$

where $\rho_\phi$ is independent of $\lambda$. One must now ask what sort of perturbation $\Delta\rho$ is to be expected. The crucial observation is that $\Delta\rho$ should also be independent of $\lambda$, otherwise one has to tune the perturbation carefully, depending on the ambient brane tension, in order to produce the requisite $\Delta\rho$. For example, one cannot produce a $\Delta\rho$ which depends on $\lambda$ from a generic scalar field theory.

The ratio $A_T^2/A_S^2$ in the braneworld satisfies

$$\frac{A_T^2}{A_S^2} = \varepsilon \left(1 - \mu F^2 Y^2 \Delta H - 2\frac{\Delta H}{H_0}\right), \tag{7.5.22}$$

which is obtained by taking the ratio of (7.5.19) and (7.4.27), where the slow-roll parameter $\varepsilon$ is defined conventionally (Maartens et al., 2000),

$$\varepsilon = \frac{2}{\kappa_4^2} F^2 \left(\frac{H_0'}{H_0}\right)^2 \frac{1}{1 + H_0^2 \ell^2}. \tag{7.5.23}$$

To complete the analysis, one only needs an expression for the tensor index $n_T$. By setting $\mathrm{d}\ln k \sim H\,\mathrm{d}t$ and replacing $t$ with the background evolution of $\phi_{\mathrm{cl}}$, one obtains

$$n_T = -2\varepsilon \left(1 - \frac{\Delta H}{H_0}\right) - 3\frac{\Delta H}{H_0} - \frac{\mathrm{d}}{\mathrm{d}\ln k}\left(\mu F^2 Y^2 \Delta H + \sqrt{\frac{\pi}{2}} Q(-k^{-1})\right). \tag{7.5.24}$$

The appropriate minimal consistency relation, in this case, should be the first-order relation $n_T = -2A_T^2/A_S^2$. One should use the first-order relation and not the next-order relation (Lidsey et al., 1997), because we do not include the next-order effects which lead to (7.1.2).

If one demands that the first-order consistency relation holds, then $F^2$, $Y^2$ $Q$ and their derivatives must be related in a particular way. Of course, one can expect to find general solutions $\Delta H$ which make this consistency relation true. But if one demands in addition that the appropriate $\Delta\rho$ is independent of $\lambda$, as we have argued above that a general matter theory should obey, then it is no longer so clear that solutions exist. Indeed, by power expanding in $\lambda$, which should be good at least in a local neighbourhood of $\lambda = 0$, one can show in the background limit where $\rho_\phi =$ constant that a solution with $\Delta\rho$ independent of $\lambda$ is not possible.

It is conceivable that solutions with $\Delta\rho$ a function of $\lambda$ exist, but such solutions are fine-tuned. In other words, it may be possible to recover the consistency relation for some choices of the matter theory, but this is no longer generic. This is the principal result of this paper: the low-order consistency relation is broken in the braneworld, in a generic manner, when perturbations away from the de Sitter background are considered.

## 7.6. Summary

In this chapter, we have applied the apparatus of five-dimensional quantum field theory to the question of gravitational perturbations in Randall–Sundrum type cosmologies. We have developed a perturbation expansion for the gravitational wave modes around the pure de Sitter case $H =$ constant which applies in the braneworld and in four dimensions. We

use this technology to calculate the power spectrum of scalars and gravitational waves as seen on the brane, or in four dimensions, and write a consistency relation in the four-dimensional case. We also show that no such consistency relation exists in the braneworld, except for fine-tuned scenarios.

The breaking of the consistency relation in the braneworld happens for a good reason, namely the presence of an extra normalization piece $Y^2$ in the four-dimensional effective action which accounts for changes in the volume of the bulk AdS slice which lies between the brane and the Cauchy horizon. This is the essential content of our simplifying approximation, which keeps only a 'breathing mode' of the bulk perturbation. Moreover, we are genuinely comparing like with like when we contrast this result in the braneworld with a four-dimensional reference geometry: in each case, the perturbation is solely to the matter component. It is important to be specific about how the perturbation occurs in the braneworld, where a perturbation to $H$ can be partitioned between ordinary matter and the dark radiation. Therefore, the extra physics which breaks the consistency relation does genuinely arise from a bulk effect, namely the change in volume between the brane and the horizon, but it is certainly not a back-reaction effect caused by scattering off Weyl curvature in the bulk. We anticipate that such back-reaction corrections would enter at a higher order in perturbation theory, but at present such refinements are out of reach of analytical treatment.

This analysis addresses a troubling feature of the braneworld model: it predicts an identical observational degeneracy in comparison with the conventional four-dimensional cosmology. We have shown, by an explicit calculation, that degeneracies of this type are not generic. Indeed, the degeneracy is broken for an open neighbourhood of models close to the de Sitter solution. Our methods do not say much about models which are distant from de Sitter space. This result is important; a complete degeneracy would hinder any attempt to observationally reconstruct the inflaton potential (Liddle and Taylor, 2002).

Our calculation relies on exploiting a technical device to calculate the tensor power spectrum in a model perturbed around a de Sitter brane carrying a single scalar field. This extends the range of models in which one knows how to solve for the spectrum of gravitational waves produced during an inflationary epoch. This is a hard problem, whose complete solution is not yet understood. Our method relies on the presence of a distinct, stable zero mode which has trivial dependence on the transverse dimension, and will not

easily generalize to full case of arbitrary time evolution on the brane, but may suggest future directions in which to proceed. One such possibility is to study the brane universe in explicitly static SAdS coordinates, where there is a holonomic timelike Killing vector $\partial/\partial T$. The graviton field equation is then independent of $T$ and becomes an ordinary differential equation similar to the Regge–Wheeler equation of black hole perturbation theory. The brane appears as a Neumann boundary condition applied to what is effectively a moving mirror, and it is possible that this framework is accessible to analytic attack. Our calculation does not yet include back reaction from other fields on the brane, and so it is not general enough (for example) to include other types of matter, or to generalize to a second order result.

CHAPTER 8

# Quantum cosmology of Randall–Sundrum type models

From the beginning it has been clear that classical cosmology—a theory of the universe wholly circumscribed by general relativity and the classical considerations of gravity—would eventually have to be replaced by a more complete theory that united gravity and the universe at large with the other fundamental forces, and above all with the uncertainty principle. Cosmologies based on brane physics, either in a manner integral to the $(3+1)$ description of our world, or at the lesser level of a mere assignment of responsibility for otherwise mysterious physics, such as inflation, to brane effects, are the first step in a process which will incorporate features of a hoped-for quantum theory of gravity in mainstream empirical cosmology.

Previous chapters have applied semiclassical approaches to the study of quantum effects in braneworlds. As a general rule, the scheme of calculation always involves solving for the background geometry in a purely classical way, according to the prescriptions of general relativity, or whatever theory of gravity is under discussion. One then introduces perturbations around this classical, background geometry and interprets these perturbations according to the standard rules of quantum mechanics. As a full description of quantum mechanical processes, however, this approach is necessarily deficient because it is presumably inadmissible to treat the background spacetime as a passive classical entity, which should instead be obliged to abide by fundamental precepts such as the uncertainty principle no less than the sundry matter fields which propagate over it.

At present there are two strongly motivated candidates for a potential theory of quantum gravity. One is the theory of quantum mechanical strings which was introduced in the late 1960s as a possible theory of the strong interaction (Green et al., 1987; Polchinski, 1998), and described in Chapter 3. Most formulations of this theory are based on the perturbative study of 'first quantized' strings, but, as we have described, recent attempts to give the theory an non-perturbative formulation have uncovered a vast and subtle network of dualities which relate string theory to gravity, gauge theory and non-commutative field

theory in appropriate limits, and in some cases allow one to follow the physics between these various sectors (Aharony et al., 2000; Horava and Witten, 1996a,b; Maldacena, 1998, 2003b). The second candidate theory, which we have not studied in this thesis, is loop quantum gravity, which is based on a fairly conservative canonical quantization of the gravitational field (Rovelli, 1997; Thiemann, 2003). However, despite the appeal of such an apparently minimal quantization, it is difficult to couple matter to loop quantum gravity, which makes the sort of phenomenology we have been considering problematic. Some progress has recently been made along these lines (Lidsey, Mulryne, Nunes, and Tavakol, 2004).

For the present, a truly fundamental, predictive theory of quantum gravity is entirely out of reach. Even if the spin 2 particle contained in string theory and exhibiting many of the familiar properties of an Einstein graviton does turn out to be the geometrodynamical mechanism nature has chosen, it is heavily premature to suggest that such a subtle, ill-understood theory as string theory could be at all predictive, or even under good calculational control. Much of the requisite technology for dealing with string theory seems presently to be missing from the mathematicians' armoury. Instead, accessible approaches to quantum gravity must necessarily be founded on approximation and error, in the hope that some genuine, albeit tentative, features of authentic quantum gravitational physics might remain visible above the muddying approximations. A relatively successful endeavour along these lines is minisuperspace quantum cosmology (Carlip, 1998; D'Eath, 1996; Dirac, 1950), based on either the canonical Schrödinger wavefunctional approach or path integral techniques (D'Eath, 1996; Hartle and Hawking, 1983; Hawking, 1984). In this chapter, constrasting with the consistent approach favouring path integrals which was adopted in earlier chapters, we shall concentrate on the Schrödinger functional approach, which has the advantage that explicit calculations can readily be carried out.

Minisuperspace models restrict the degrees of freedom which occur in full quantum gravity to a small number, preferably finite, which describe a highly symmetrical geometry. One might hope that such a drastic reduction in the number of degrees of freedom would eliminate, or at least render accessible, many of the conceptual and technical issues which conspire to render full quantum gravity intractible. This does indeed occur, but much of the subtlety and dynamical richness of the theory is lost, so that it is not clear how much of

the genuine quantum dynamics is visible and the conclusions of the theory must be hedged around with technicalities and caveats (Carlip, 1998).

Aside from the general desire to put brane cosmology on some sort of secure theoretical footing, by requiring it to abide by the stipulations and principles of quantum mechanics in the same way as all other theories of contemporary physics, there are other reasons for embarking on a study of quantum brane cosmology. The lengthy algebraic calculations of previous chapters have highlighted a quite general malaise of string compactifications: the low energy effective field theories can be difficult to handle. Although the field theory approach has led to some success, as we have described, in the prediction of the spectrum of tensor modes prouced from an epoch of early universe inflation (Gorbunov et al., 2001; Langlois et al., 2000) – and also in the discovery of degeneracies in the predicted observables of the theory (Huey and Lidsey, 2001, 2002; Liddle and Taylor, 2002), as described in Chapter 7 – theoretical progress in this direction has effectively been hampered by an inability to solve for the graviton field modes in the transverse dimension when the braneworld is not of the fibre bundle structure outlined in Chapter 5. Given this impasse, it is natural to seek alternative represenations of the physics, which might suggest more manageable calculational techniques.

The quantum cosmology of brane universes has already received some attention in the literature, from a rather different perspective. Usually, the use of quantum cosmological techniques is reserved for the study of the 'creation' of the universe, a tunelling of a quantum FRW metric from zero to finite radius, or similar exotic effects. Because the gravitational field has a gauge invariance corresponding to reparametrization of the time (Dirac, 1950; Higgs, 1958), the quantum representation does not involve the timelike coordinate, but rather only the spatial geometry. Therefore, this representation might prove easier to solve than the explicit five-dimensional wave equation. Although quantum cosmological methods might not give the same information as the low energy field theory, its predictions would still be profitable. For example, an estimate of the temperature inhomogeneities in the cosmic microwave background can be given in four-dimensional models using this technique (Halliwell and Hawking, 1985). This approach would involve trading one kind of difficulty, the solution of a partial differential equation in curved coordinates, for another, the solution of the constraints of quantized general relativity.

The origin of brane physics in gauge theory, string theory and supergravity equally encourages many diverse avenues of investigation. Quantum cosmology is possible using a variety of techniques familiar from four dimensions: the canonical Wheeler–de Witt formalism (Biswas, Mukherji, and Pal, 2004; Koyama and Soda, 2000; Sanyal, 2003; Seahra, Sepangi, and Ponce de Leon, 2003); tunelling instantons (Copeland, Gray, and Saffin, 2000; Garriga and Sasaki, 2000; Gray and Copeland, 2001), and other instanton-like solutions (Cordero and Rojas, 2003; Cordero and Vilenkin, 2002; Gregory and Padilla, 2002; Ida, Shiromizu, and Ochiai, 2002); and the AdS/CFT correspondence (Elizalde, Nojiri, Odintsov, and Ogushi, 2003; Nojiri and Odintsov, 2001, 2002, 2003). In the present context, we work with the canonical formalism. The effective Wheeler–de Witt equation on the brane, in a four-dimensional minisuperspace approximation, was first written down by Koyama and Soda (2000), who used the resulting transition probabilities to study quantum tunnelling of the universe from nothing to a finite radius. (Such a possibility had already been considered from a different perspective in Garriga and Sasaki (2000).) This tunelling can, under suitable hypotheses, be identified with a creation event for the universe. In this and all subsequent work, the analysis was based on a four-dimensional effective action, which was obtained by a suitable compactification and truncation of the full five-dimensional Randall–Sundrum model. Such a truncation reduces the model to a single dimension, so one is effectively dealing with quantum mechanics rather than field theory, and the model can be expected to be at least approximately solvable. Despite the apparent limitations this procedure implies, several interesting questions can usefully be addressed within this framework.

The basic conclusions resulting from this work were that the braneworld Wheeler–de Witt equation, given the various approximation which were made, and provided one was dealing with an exactly AdS bulk, should coincide with the Wheeler–de Witt equation of conventional cosmology (Koyama and Soda, 2000) in the case of zero bulk Schwarzschild mass. Moreover, the matter sector decoupled and could be considered separately. However, it has long been known the the four-dimensional gravity induced on the brane is not quite Einstein gravity but a modification which reduces to Einstein gravity at low energy (Hawking et al., 2000; Perez-Victoria, 2001) but couples to matter differently at high energy. This high energy modification arises because of the freedom of gravity to explore

the extra transverse dimensions, and one might speculate that a signature of this freedom should appear in the quantum framework. We shall see later that this does indeed occur.

Several years later, Biswas et al. (2004) used a very similar formalism to study the resolution of cosmological singularities within the brane model, and showed that generically the wavefunction of the universe can be chosen to obey de Witt boundary conditions at the singularity, that is,

$$\Psi(R) \to 0 \quad \text{at } R = 0, \tag{8.0.1}$$

where $R$ is the scale factor of the universe. This choice of boundary conditions provides some encouragement for the idea that the cosmological singularity is accessible to investigation, since the universe collapses to zero size only with probability zero. On the basis of this and similar calculations, the authors of Biswas et al. (2004) proposed that the quantum brane universe might be stabilized around the Planck scale, and never decrease to zero volume. Quantum cosmology with AdS spaces of different radii on either side of the brane have also been considered (Seahra, 2003; Seahra et al., 2003).

Quantum cosmological effects are speculative, but important. In particular, they provide another test of the ability of brane-based cosmological scenarios to replicate the standard model. Had the braneworld failed this test in a way that could not be reconciled with observation or theoretical analysis of the early universe, it would have been a heavy blow to any attempt to import concepts from string theory into standard cosmology by this route. Quantum cosmology, viewed in this way, properly belongs to the phenomenological effort which constitutes this thesis. The early work on the Wheeler–de Witt equation shows, reassuringly, that the general scheme of quantum cosmology which has been erected in four dimensions can be carried over more or less intact to the case of the braneworld, although there are some surprises. Most notably, as we have observed, the approximations of Koyama and Soda (2000) imply that the Wheeler–de Witt equation misses the quadratic corrections which appear in the brane Friedmann equation (Binetruy et al., 2000a,b). This is rather unexpected, because such corrections non-trivially modify the early evolution of the universe.

The work described in this chapter differs from the foregoing analyses, because it is not based on a four-dimensional effective action. Instead, the full five-dimensional Randall–Sundrum action is retained, which, when restricted to four-dimensional homogeneous,

isotropic cosmologies is equivalent to the action for a particular $(1+1)$-dimensional diffeomorphism invariant field theory. This field theory is not exactly solvable (indeed, it would have been most remarkable if it were), but an approximate solution for the Schrödinger wavefunctional can be given for certain field configurations provided the AdS radius is large. In addition, one can study the boundary theory on the brane almost independently of the bulk dynamics. Unfortunately, it turns out that difficulties in the solution of the quantum constraints almost certainly mean that it is not practical to attempt to study bulk quantum perturbations by this route. As a by-product, we obtain an alternative way to study the Wheeler–de Witt equation on the brane without first constructing an effective action.

In Section 8.1 we outline the Schrödinger treatment of general relativity in four dimensions. This formalism is applied to Randall–Sundrum minisuperspaces in Section 8.2, for which we derive the Hamiltonian form of the action principle. After this, it is easy to verify that our formalism reproduces the familiar results of Chapter 5. The theory we arrive at is, in the gravitational sector, a theory of real-valued maps $v$ defined on an interval of the real line. One must decide, among the various possible choices, which functions are to contribute to the path integral. In the Schrödinger representation, where one chooses the quantum Hilbert space to be a suitable space of functionals $\Psi[v]$, this is equivalent to deciding the space of functions $v$ on which the wavefunction should have support. The coordinate choice is important here: we discuss the role of coordinate horizons in the bulk in Section 8.2.2, and give a short, new derivation of the coordinate transformation which takes Gaussian normal coordinate to global SAdS coordinates.

In Section 8.3 we introduce a simple model of scalar matter on the brane, which is conformally coupled to baneworld gravity. The wavefunctional decomposes into several coupled sectors: a matter sector, which can be treated independently Koyama and Soda (2000); a boundary term, which can be interpreted as the brane wavefunction; and a bulk piece, which is described by an auxiliary $(0+1)$-dimensional quantum theory coupled to the boundary sector via an integral related to the Airy function. The matter couplings only enter the boundary term, which can therefore be investigated separately. We show explicitly that the boundary wavefunction can obey the de Witt boundary conditions derived in Biswas et al. (2004). However, we find extra couplings to the matter theory that constitute corrections to the simple four-dimensional GR wavefunction. We draw

comparisons with the four-dimensional case in Section 8.3.3. In Section 8.4, we investigate in greater detail the description of the bulk sector of the wavefunction in terms of an auxiliary quantum theory. Where our approximations are valid, this theory is close to its infinite mass or strongly-coupled limit, with the anti-de Sitter cosmological constant playing the role of the deformation parameter $\hbar$.

## 8.1. Quantum cosmology in $(3+1)$-dimensions

In the case where spacetime can be foliated into spacelike slices $\Sigma$ (and therefore, according to a classical theorem of Geroch, necessarily has topology $\mathbf{R} \times \Sigma$), the ADM decomposition of the metric into a three-geometry $h_{ij}$, a lapse function $N$ and a shift vector $N^i$ is

$$\mathrm{d}s^2 = h_{ij}(\mathrm{d}x^i + N^i\,\mathrm{d}t)(\mathrm{d}x^j + N^j\,\mathrm{d}t) - (N\,\mathrm{d}t)^2. \tag{8.1.1}$$

The three-geometry $h_{ij}$ gives the intrinsic geometry on the slices $\Sigma$, whereas $N$ and $N^i$ describe, respectively, how much proper time elapses in moving between members of the foliation, and the displacement one suffers in doing so. We begin with the Einstein–Hilbert action,

$$S_{EH} = \frac{1}{2\kappa^2}\int_M \mathrm{d}^4x\,\sqrt{-g}\,R \tag{8.1.2}$$

where $R$ is the Ricci scalar of the four-geometry and $g = \det g_{ab}$. This can be rewritten in terms of ADM quantities by applying the Gauss–Codacci equation (Hawking and Ellis, 1973),

$$\overset{3}{R} = R + 2R_{ab}n^a n^b - (\mathrm{Tr}\,K)^2 + \mathrm{Tr}\,K^2 \tag{8.1.3}$$

where $\overset{3}{R}$ is the curvature scalar on slices $\Sigma$, and $K_{ab} = \nabla_a n_b$ is the second fundamental form, given that the unit vector $n_b$ is everywhere normal to $\Sigma$. To find $R_{ab}n^a n^b$, one uses the Ricci identity to commute covariant derivatives over $n^a$,

$$\nabla_a\nabla_b n_c - \nabla_b\nabla_a n_c = R_{abcd}n^d, \quad \text{so} \quad R_{bd}n^d = \nabla_a\nabla_b n^a - \nabla_b\nabla_a n^a. \tag{8.1.4}$$

In terms of ADM quantities, the volume Jacobian $\sqrt{-g}$ can be written $N\sqrt{h}$ (with, following convention, $h = \det h_{ab}$), so the Einstein–Hilbert action is entirely equivalent to

$$S_{EH} = \frac{1}{2\kappa^2}\int \mathrm{d}t\,\mathrm{d}^3x\, N\sqrt{h}\left[\overset{3}{R} + (\mathrm{Tr}\,K)^2 - \mathrm{Tr}\,K^2 - \right.$$
$$\left. 2\nabla_a(n^b\nabla_b n^a) + 2\nabla_a n^b\nabla_b n^a + 2\nabla_b(n^b\nabla_a n^a) - 2\nabla_b n^b\nabla_a n^a\right]. \tag{8.1.5}$$

The two total derivatives can be straightforwardly dealt with by converting them to surface integrals via Stokes' theorem,

$$\int_M N\sqrt{h}\,\nabla_a(n^b\nabla_b n^a) = \int_{\partial M}\sqrt{h}\,n_a n^b\nabla b n^a = 0 \tag{8.1.6}$$

$$\int_M N\sqrt{h}\,\nabla_b(n^b\nabla_a n^a) = \int_{\partial M}\sqrt{h}\,n_b n^b\nabla_a n^a = -\int_{\partial M}\sqrt{h}\,\mathrm{Tr}\,K, \tag{8.1.7}$$

where the bounding hypersurface has been taken to be timelike (that is, $n_b n^b = -1$), and we have used the fact that $\mathrm{Tr}\,\nabla_a n_b = \mathrm{Tr}\,K_{ij}$. The first surface integral is zero in virtue of the fact that $n_a\nabla_b n^a \propto \nabla_b(n_a n^a) = 0$.

The remaining terms involving derivatives are

$$\nabla_a n^b\nabla_b n^a = \mathrm{Tr}\,K^2, \quad \text{and} \quad \nabla_b n^b\nabla_a n^a = (\mathrm{Tr}\,K)^2, \tag{8.1.8}$$

which follow on expanding $h^{ab}$ in terms of $g^{ab}$ and $n^a$, and once again using the identity $n_a\nabla_b n^a = 0$. Thus,

$$S_{EH} + \frac{1}{\kappa^2}\int_{\partial M}\sqrt{h}\,\mathrm{Tr}\,K = \frac{1}{2\kappa^2}\int_M N\sqrt{h}\mathrm{d}^3x\left[\overset{3}{R} - (\mathrm{Tr}\,K)^2 + \mathrm{Tr}\,K^2\right]. \tag{8.1.9}$$

The quantity appearing on the right-hand side of (8.1.9) has the property that it is additive over cobordant spacetime regions, that is, spacetime regions sharing a common spacelike boundary, *provided* that the metric is continuous. This property is crucial in the path integral (D'Eath, 1996), where functional integrals are typically defined by slicing the integration region into strips, and assuming that the total action can be evaluated by summing the action in each strip individually. The value of the path at each strip boundary is integrated over independently, subject only to the condition that the path is continuous, so a discontinuity in the first derivative is present. The surface term on the left-hand side of (8.1.9) serves, in effect, to eliminate troublesome first derivatives on $\partial M$; the entirety

of the left-hand side of (8.1.9) is called the modified gravitational action, $S_{\rm mod}$, and the surface term is known as the Gibbons–Hawking term.

**8.1.1. The gravitational Hamiltonian.** To cast the action in Hamiltonian form, one defines the momentum canonical to $h_{ij}$ as

$$\frac{1}{2\kappa^2}\pi^{ij} = \frac{\partial L_{\rm mod}}{\partial \dot{h}_{ij}} = -\frac{1}{2\kappa^2} N\sqrt{h}\left(2\,\mathrm{Tr}\,K\frac{\partial K}{\partial \dot{h}_{ij}} - 2K^{ij}\frac{\partial K_{ij}}{\partial \dot{h}_{ij}}\right). \tag{8.1.10}$$

The factor of $(2\kappa^2)^{-1}$ is introduced for future convenience. To evaluate this, it is only necessary to know the dependence of $K_{ij}$ on $\dot{h}_{ij}$, since no other quantity in the modified action depends on $\dot{h}_{ij}$. Writing the extrinsic curvature as

$$K_{ab} = \nabla_a n_b, \quad \text{so} \quad K_{ij} = -N\Gamma^0_{ij} \quad (\text{since } n_a = (N,0,0,0)), \tag{8.1.11}$$

one has

$$\begin{aligned} K_{ij} &= -Ng^{00}\Gamma_{0ij} - Ng^{0m}\Gamma_{mij} \\ &= \frac{1}{2N}\left(\partial_j g_{0i} + \partial_i g_{0j} - \partial_0 g_{ij} - 2N^m \overset{3}{\Gamma}_{mij}\right) \\ &= \frac{1}{N}\left(N_{(i|j)} - \frac{1}{2}\frac{\partial h_{ij}}{\partial t}\right), \end{aligned} \tag{8.1.12}$$

where $|$ denotes the covariant derivative compatible with the three-geometry $h_{ij}$ and connexion $\overset{3}{\Gamma}$. Thus,

$$\frac{\partial K_{rs}}{\partial \dot{h}_{ij}} = -\frac{1}{2N}\delta^r_i\delta^j_s \quad \text{and} \quad \frac{\partial K}{\partial \dot{h}_{ij}} = -\frac{1}{2N}h^{ij} \quad \text{implies} \quad \pi^{ij} = -\sqrt{h}(K^{ij} - h^{ij}\,\mathrm{Tr}\,K). \tag{8.1.13}$$

To obtain the Hamiltonian, one rewrites $L_{\rm mod}$ in terms of $\pi^{ij}$ and $h_{ij}$, eliminating derivatives $\dot{h}_{ij}$ of the three-geometry. This proceeds in several stages. First, it is easy to show that $\mathrm{Tr}\,\pi^2$ satisfies the identity

$$\mathrm{Tr}\,\pi^2 = h\left[\mathrm{Tr}\,K^2 + (\mathrm{Tr}\,K)^2\right], \tag{8.1.14}$$

so the term $(\mathrm{Tr}\,K)^2$ term can be eliminated at once from the Lagrangian:

$$2\kappa^2 L_{\rm mod} = N\sqrt{h}\,\overset{3}{R} + 2N\sqrt{h}\,\mathrm{Tr}\,K^2 - \frac{N}{\sqrt{h}}\,\mathrm{Tr}\,\pi^2. \tag{8.1.15}$$

To finish the job requires an expression for $\mathrm{Tr}\,K^2 = K^{ij}K_{ij}$, which is easy to obtain by rewriting $\pi^{ij}$ for $K^{ij}$ and taking the product with $K_{ij}$ in the explicit form involving $N_i$

and $\dot{h}_{ij}$. One has

$$N\sqrt{h}\ \mathrm{Tr}\, K^2 = (-\pi^{ij} + \sqrt{h}\, h^{ij}\, \mathrm{Tr}\, K)(N_{(i|j)} - \frac{1}{2}\dot{h}_{ij}) \tag{8.1.16}$$

$$= -\pi^{ij} N_{i|j} + \frac{1}{2}\pi^{ij}\dot{h}_{ij} + \sqrt{h}\ \mathrm{Tr}\, K\, h^{ij} N_{i|j} \tag{8.1.17}$$

where we have used the fact that $h^{ij}\dot{h}_{ij} = 0$. Substituting into $L_{\mathrm{mod}}$, we find

$$2\kappa^2 L_{\mathrm{mod}} = N\sqrt{h}\ \overset{3}{R} - 2\pi^{ij} N_{i|j} + \pi^{ij}\dot{h}_{ij} + \sqrt{h}\ \mathrm{Tr}\, K\, h^{ij} N_{i|j}. \tag{8.1.18}$$

The only troublesome term is the last, involving a trace over $N_{i|j}$. This has to be eliminated in order to arrive at the correct Hamiltonian, since it involves the explicit trace of the extrinsic curvature, and therefore $\dot{h}_{ij}$. In fact, $\mathrm{Tr}\, K$ and $\mathrm{Tr}\, N_{i|j}$ are proportional:

$$2NK_{ij} = 2N_{(i|j)} - \dot{h}_{ij} \quad \text{implies} \quad 2N\,\mathrm{Tr}\, K = 2\,\mathrm{Tr}\, N_{i|j}, \tag{8.1.19}$$

since, as we have previously pointed out $h^{ij}\dot{h}_{ij} = 0$. Therefore this term becomes $2N\sqrt{h}\,(\mathrm{Tr}\, K)^2$. To finally eliminate the $\mathrm{Tr}\, K$, one simply traces over $\pi^{ij}$ to find $\mathrm{Tr}\,\pi = 2\sqrt{h}\ \mathrm{Tr}\, K$. Therefore,

$$2\kappa^2 L_{\mathrm{mod}} = \pi^{ij}\dot{h}_{ij} - N\left[\frac{1}{\sqrt{h}}(\mathrm{Tr}\,\pi^2 - \frac{1}{2}(\mathrm{Tr}\,\pi)^2) - \sqrt{h}\overset{3}{R}\right] - 2\pi^{ij} N_{i|j}. \tag{8.1.20}$$

It is more convenient to rewrite the term involving $N_{i|j}$ as

$$-2\pi^{ij} N_{i|j} = -2(\pi^{ij} N_i)_{|j} + 2\pi^{ij}{}_{|j} N_i, \tag{8.1.21}$$

and the resulting surface term can be safely discarded. Defining the Hamiltonian, sometimes called the super-Hamiltonian in the context, by $H = \pi^{ij}\dot{h}_{ij} - 2\kappa^2 L_{\mathrm{mod}}$, we have

$$H = \int \mathrm{d}^3x\ N\mathcal{H} + N_i\mathcal{H}^i \tag{8.1.22}$$

where

$$\mathcal{H} = \frac{1}{\sqrt{h}}\left(\mathrm{Tr}\,\pi^2 - \frac{1}{2}(\mathrm{Tr}\,\pi)^2\right) - \sqrt{h}\ \overset{3}{R} \tag{8.1.23}$$

$$\mathcal{H}^i = -2\pi^{ij}{}_{|j}. \tag{8.1.24}$$

**8.1.2. The Wheeler–de Witt equation.** The variational equations with respect to $N$ and $N_i$ show that, classically, the three-geometry obeys the constraints

$$\mathcal{H} = 0 \quad \text{and} \quad \mathcal{H}^i = 0. \tag{8.1.25}$$

One passes to the quantum theory by imposing the canonical commutation relations

$$[h_{ij}, (2\kappa^2)^{-1}\pi^{rs}] = \mathrm{i}\hbar\delta^r_{(i}\delta^s_{j)}. \tag{8.1.26}$$

In the coordinate representation, these relations can be solved by taking $h_{ij}$ as an operator which multiplies its argument by $h_{ij}(x)$ and setting

$$\pi^{rs}(x) = -2\kappa^2\mathrm{i}\hbar\frac{\delta}{\delta h_{rs}(x)}. \tag{8.1.27}$$

The classical constraints then become weak operator constraints on states in the Hilbert space, by demanding that $\widehat{\mathcal{H}}\Psi[h_{ij}] = 0$ (and an equivalent constraint for $\mathcal{H}^i$) for Schrödinger functionals $\Psi[h_{ij}]$. To do this conveniently, one rewrites $\mathcal{H}$ to remove explicit trace operators,

$$G_{ijkl}\pi^{ij}\pi^{kl} - \sqrt{h}\overset{3}{R} = 0, \tag{8.1.28}$$

where the de Witt metric $G_{ijkl}$ is defined by

$$G_{ijkl} = \frac{1}{2}\frac{1}{\sqrt{h}}\left(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}\right). \tag{8.1.29}$$

On quantization, this goes over to

$$\left[-(2\kappa^2\hbar^2)G_{ijkl}\frac{\delta}{\delta h_{ij}(x)}\frac{\delta}{\delta h_{kl}(x)} - \sqrt{h}\overset{3}{R}\right]\Psi[h_{ij}] = 0. \tag{8.1.30}$$

This is known as the Wheeler–de Witt equation. A similar but very much simpler equation holds for each of the spatial constraints $\mathcal{H}^i = 0$.

## 8.2. Quantum Randall–Sundrum universes

**8.2.1. The gravitational Hamiltonian.** We begin with the Lagrangian formulation of the Randall–Sundrum model. In order to turn this into a properly defined canonical quantum theory, it is necessary to make the transition to a phase space description, in terms of which the fundamental quantities are a Hamiltonian and Hamilton's action principle. For the present, we are concerned only with the pure gravitational dynamics. The case where matter resides on the brane will be considered later (Section 8.3).

The action is $S = (2\kappa^2)^{-1}\int \mathrm{d}^5x\,(R+2\Lambda)$ which is appropriate for Einstein gravity with a negative cosmological constant. The vacuum extrema are topologically anti-de Sitter

spaces. One now chooses coordinates so that the metric is described by the Randall–Sundrum line element,

$$\mathrm{d}s^2 = -n^2(t,y)\,\mathrm{d}t^2 + a^2(t,y)\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j + \mathrm{d}y^2, \tag{8.2.1}$$

where the fields $n$ and $a$ are the elementary fields of the model, and $\gamma_{ij}$ is any maximally symmetric three-geometry. The most interesting cases occur where $\gamma$ is closed or flat. For reasons associated with the addition of conformally coupled scalar matter, to be discussed in Section 8.3 below, we choose the closed model for to be definite and indicate differences with the flat case where they arise. The modified Einstein action is

$$S_{\mathrm{mod}} = \frac{V}{2\kappa^2}\int_M \mathrm{d}t\,\mathrm{d}y\;\mathcal{L} - \frac{V}{2\kappa^2}\int_{\partial M}\mathrm{d}t\;2\lambda n a^3 + \text{Gibbons–Hawking terms}, \tag{8.2.2}$$

where the bulk Lagrangian density $\mathcal{L}$ satisfies

$$\mathcal{L} = 6na - 6naa'^2 - 6na^2a'' - 6a^2a'n' - 2a^3n'' - 6\frac{a^2\dot{a}\dot{n}}{n^2} + 6\frac{a\dot{a}^2}{n} + 6\frac{a^2\ddot{a}}{n} + 2\Lambda n a^3. \tag{8.2.3}$$

$M$ is a two-dimensional $\mathbf{Z}_2$-symmetric manifold, and $\lambda$ is the intrinsic tension on the brane $\partial M$, which is fixed at $y = 0$. The volume $V = \int \mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi\, r^2(1-r^2)^{-1/2}\sin\theta$ is the spatial volume appropriate for a closed universe. One can derive the Gibbons–Hawking terms either from integration by parts in the action or directly from their definition in terms of the second fundamental form $K_{ab}$. They are

$$-\frac{V}{\kappa^2}\int_{t=\text{const.}}\mathrm{d}y\;\frac{3}{n}a^2\dot{a} \quad\text{and}\quad \frac{V}{\kappa^2}\int_{y=\text{const.}}\mathrm{d}t\;(n'a^3 + 3na^2a'). \tag{8.2.4}$$

After reshuffling terms between the boundary and bulk action in order to eliminate troublesome derivatives of $n$, one is left with the reduced action

$$S_{\mathrm{mod}} = \frac{\alpha}{2}\int_M \mathrm{d}t\,\mathrm{d}y\;\left(-\frac{a\dot{a}^2}{n} + na - naa'^2 - na^2a''\right) - \frac{\alpha}{2}\int_{\partial M}\mathrm{d}t\left(\frac{\lambda}{3}na^3 + na^2[a']^+_-\right) \tag{8.2.5}$$

where the jump $[z]^+_-$ in some quantity $z$ across $\partial M$ is defined by

$$[z]^+_- = \lim_{y\to 0^+} z - \lim_{y\to 0^-} z, \tag{8.2.6}$$

and $\alpha = 6V/\kappa^2$. Although it is necessary to use the Gibbons–Hawking term which arises from temporal slicing to eliminate the second order derivative $\ddot{a}$, in order to produce a well-defined variational principle which works in the path integral, this is not necessary in the case of the $y$-derivative $a''$. The absence of $\ddot{a}$ is required to ensure additivity of the action in the temporal-slicing which is used to define the path integral, but there is no such

requirement for $a''$. Indeed, a considerable gain in convenience results from leaving the $a''$ term in the bulk action and dealing with the boundary contribution separately. The term $na$ in $S_{\rm mod}$ is absent for a flat universe. In Hamiltonian form, one has

$$S_{\rm mod} = \int {\rm d}t\,{\rm d}y\ \left[\pi\dot{a} - n\left(-\frac{1}{2}\frac{\pi^2}{\alpha a} - \frac{\alpha}{2}a + \frac{\alpha}{2}aa'^2 + \frac{\alpha}{2}a^2a'' - \frac{\alpha}{2}\frac{\Lambda}{3}a^3\right)\right] - \frac{\alpha}{2}\int {\rm d}t\ na^2\left([a']^+_- + \frac{\lambda}{3}a\right). \tag{8.2.7}$$

As a check on the correctness of this reduction, it is easy to verify that the familiar (Binetruy et al., 2000a,b) field equations for the Randall–Sundrum model are recovered. The bulk $\pi$ variation merely gives back a definition for the momentum $\pi$,

$$\pi = -\frac{\alpha a\dot{a}}{n}, \tag{8.2.8}$$

whereas the bulk $n$ variation gives the momentum constraint,

$$-\frac{\pi^2}{2\alpha a} - \frac{\alpha}{2}a + \frac{\alpha}{2}aa'^2 + \frac{\alpha}{2}a^2a'' - \frac{\alpha}{2}\frac{\Lambda}{3}a^3 = 0. \tag{8.2.9}$$

The variation of $n$ on $\partial M$ gives a jump condition on $a'$,

$$[a']^+_- = -\frac{\lambda}{3}a|_{y=0}. \tag{8.2.10}$$

This is the well-known Lanczos–Israel matching condition in the brane context (Binetruy et al., 2000a). Similarly, the bulk $a$ variation gives a evolution equation for $\pi$,

$$\dot{\pi} + \frac{n\pi^2}{2\alpha a^2} - \frac{\alpha n}{2} + \frac{\alpha n}{2}a'^2 + \frac{\alpha n''}{2}a^2 + \alpha n'aa' + \alpha naa'' - \frac{\alpha}{2}\Lambda na^2 = 0. \tag{8.2.11}$$

Using the definition of $\pi$, it can be shown that this evolution equation is equivalent to the $(i,j)$ Einstein equation which would be derived from the Randall–Sundrum line element,

$$\frac{1}{n^2}\left(\frac{\dot{a}}{a}\right) + \frac{2}{n^2}\frac{\ddot{a}}{a} - \frac{2}{n^2}\frac{\dot{n}}{n}\frac{\dot{a}}{a} - \left(\frac{a'}{a}\right)^2 - 2\frac{n'}{n}\frac{a'}{a} - \frac{n''}{n} - 2\frac{a''}{a} + \frac{1}{a^2} + \Lambda = 0. \tag{8.2.12}$$

The term involving only $a^{-2}$ is absent for a flat model. The momentum constraint itself is equivalent to the time–time Einstein equation,

$$\frac{1}{n^2}\frac{\dot{a}^2}{a^2} - \left(\frac{a'}{a}\right)^2 - \frac{a''}{a} + \frac{1}{a^2} + \frac{\Lambda}{3} = 0. \tag{8.2.13}$$

The field $n$ is a Lagrange multiplier associated with a gauge degree of freedom, and can be chosen more-or-less arbitrarily during the classical evolution, subject to the restriction $[n']^+_-/n = [a']^+_-/$ derived from the boundary variation of $a$. If one makes the choice

$n = \dot{a}\beta^{-1}$ (Binetruy et al., 2000b), where $\beta$ is a function only of $t$, then the momentum constraint admits an immediate first integral,

$$\frac{\beta^2}{a^2} + \frac{1}{a^2} + \frac{\Lambda}{6} - \left(\frac{a'}{a}\right)^2 = \frac{\mathcal{C}}{a^4}, \tag{8.2.14}$$

where $\mathcal{C}$ is an arbitrary constant of integration (see below). Making use of the jump condition and evaluating on $\partial M$, having chosen $\beta$ to equal to scale factor on the brane, yields the brane Friedmann equation

$$H^2 = \frac{\Lambda_4}{3} - \frac{1}{a_{\rm b}^2} + \frac{\mathcal{C}}{a_{\rm b}^4} \quad \text{where} \quad \Lambda_4 = \frac{1}{2}\left(\frac{\lambda^2}{6} - \Lambda\right). \tag{8.2.15}$$

It was not necessary to solve for the bulk geometry to obtain this result (compare Gubser (2001)). In the absence of matter and $\mathcal{C}$, this is identically the Friedmann equation for a closed four-dimensional universe, or a flat universe if the term $a_{\rm b}^{-2}$ is removed.[1]

**8.2.2. Black hole masses in the bulk.** Anti-de Sitter space is not the unique solution to the Einstein equations with negative cosmological constant. Instead, one may allow black holes immersed in anti-de Sitter space, so called Schwarzschild-anti de Sitter space or SAdS. This possibility arises via the constant of integration $\mathcal{C}$, which can be shown to be induced by a Schwarzschild-like mass in the bulk. Quantum cosmology in the presence of bulk mass was considered in Koyama and Soda (2000); Seahra et al. (2003), whereas the general relation between the Gaussian normal coordinates in which the present discussion is framed, and global SAdS coordinates, was given by Mukohyama et al. (2000).

There several reasons for being careful about the question of global versus local coordinates. Most importantly, one must have a clear idea which class of metrics should contribute to the path integral, or, in the Scrödinger representation, over which metrics the wavefunctional should have support. One must also understand the classical geometry in order to have any hope of approaching a consistent quantum field theory, and, in particular, one must understand the role of horizons and boundaries. There may be more than one consistent quantization of a given theory. This is of particular relevance in AdS, and spaces asymptotic to it, where the boundary conditions one applies to fields at the brane

[1]The observation that the quantum version of this theory (Koyama and Soda, 2000) coincides with the quantum four-dimensional universe is rather natural in this context. In other words, the classical theory one has obtained in the absence of matter on the brane is identical with the four-dimensional result so it is not surprising, naïvely speaking, that the quantum theory corresponding to it is the same. It is only when addition matter is added that the differences involved in braneworld gravity become manifest.

and at the so-called Cauchy horizon in the bulk are crucial in picking out a set of classical solutions to the field equations from which to build the quantum theory (Breitenlohner and Freedman, 1982a,b; Mezincescu and Townsend, 1985).

The coodinate transformation which relates Gaussian normal coordinates to global Schwarzschild coordinates was derived in Bowcock et al. (2000); Mukohyama et al. (2000). This procedure is rather complicated. Here, we present a simpler derivation. The new step in this approach involves finding an appropriate timelike Killing vector on which base the SAdS coordinate construction.[2] Denote this Killing vector $\partial/\partial T$. The condition that $\partial/\partial T$ be Killing is that the Lie-dragging of the metric along the flow generated by $\partial/\partial T$ is zero, or, if we suppose

$$\frac{\partial}{\partial T} = r(t,y)\frac{\partial}{\partial t} + s(t,y)\frac{\partial}{\partial y}, \tag{8.2.16}$$

one obtains

$$s' = 0, \quad \dot{a}r + a's = 0, \quad \dot{s} = n^2 r', \quad \text{and} \quad r\dot{n} + sn' + n\dot{r} = 0, \tag{8.2.17}$$

where an overdot represents a derivative with respect to $t$, whereas a prime $'$ denotes a $y$-derivative. This is just the condition $(\pounds_{\partial/\partial T})g_{ab} = 0$ written out in components, where $\pounds_{\mathbf{X}}$ is the Lie derivative along $\mathbf{X}$ and $g_{ab}$ is taken to be the metric (8.2.1). The first of these implies that $s$ is independent of $y$, so the others can be written as ordinary differential equations for $s(t)$. We obtain

$$r(t,y) = -\frac{a'(t,y)}{\dot{a}(t,y)}s(t) \quad \text{and so} \quad \frac{\mathrm{d}\log s}{\mathrm{d}t} = -n^2(t,y)\left(\frac{a'(t,y)}{\dot{a}(t,y)}\right)' = -n^2\left(\frac{a''}{\dot{a}} - \frac{a'}{\dot{a}}\frac{\dot{a}'}{\dot{a}}\right). \tag{8.2.18}$$

Using the Einstein field equations, and supposing that $n(t,y) = \dot{a}(t,y)/\dot{a}_{\mathrm{b}}(t)$ (where, as usual, $a_{\mathrm{b}} = a(y=0)$, and is a choice that can always be made under mild hypotheses

[2]The spacetime described by (8.2.1) has Killing vectors associated with the spatial three-metric $\gamma_{ij}$, which are not affected by the embedding in AdS. The presence of the brane breaks the isometry corresponding to bulk translations, which would otherwise be generated by $\partial/\partial y$. This is not trivial: for example, there is no conserved bulk momentum, corresponding to a Noether charge, because the isometry which would generate it is absent. In addition, cosmological evolution breaks time translations generated by $\partial/\partial t$ on the brane, but since the bulk geometry remains AdS (or SAdS), one expects a residual timelike Killing vector associated with isometries of the bulk space. We can anticipate that the isometry might be broken at the brane, or the Cauchy horizon, corresponding to boundaries of the SAdS patch, both of which will in fact turn out to be the case.

about the behaviour of the bulk, in particular that $T_{ty}$ vanishes), one can show that the right-hand side is

$$-n^2\left(\frac{a''}{\dot{a}}-\frac{a'}{\dot{a}}\frac{\dot{a}'}{\dot{a}}\right)=\frac{\mathrm{d}\log\dot{a}_{\mathrm{b}}}{\mathrm{d}t}+\delta_{\mathrm{D}}\text{-function terms},\tag{8.2.19}$$

where the $\delta_{\mathrm{D}}$-function terms spoil the isometry at the brane. The $s$-equation can be integrated at once in the bulk to give $s=\dot{a}_{\mathrm{b}}$. These expressions hold regardless of the value of the bulk cosmological constant, the effective four-dimensional cosmological constant, or any putative black hole mass. Since the overall scale of $\partial/\partial T$ is not important for our applications, we can write

$$\frac{\partial}{\partial T}=\dot{a}_{\mathrm{b}}\left(\frac{\partial}{\partial y}-\frac{a'}{\dot{a}}\frac{\partial}{\partial t}\right).\tag{8.2.20}$$

$T$ will be the SAdS time coordinate. To finish the construction only requires obtaining another holonomic basis vector. Frobenius' Theorem shows that a basis is coordinate induced if and only if all Lie brackets of the basis elements among themselves vanish. In the present case, it is clear that $\partial/\partial T$ commutes with the three-geometry vectors $\partial/\partial x^i$, but that neither $\partial/\partial t$ nor $\partial/\partial y$ commute with $\partial/\partial T$. To resolve this, introduce a new coordinate $r=a$, with corresponding one-form

$$\mathrm{d}r=\dot{a}\,\mathrm{d}t+a'\,\mathrm{d}y.\tag{8.2.21}$$

The one-form $\mathrm{d}r$ is dual to a vector $\partial/\partial\rho$ which satisfies

$$\frac{\partial}{\partial\rho}=-\frac{\dot{a}_{\mathrm{b}}^2}{\dot{a}}\frac{\partial}{\partial t}+a'\frac{\partial}{\partial y}.\tag{8.2.22}$$

The vectors $\partial/\partial\rho$ and $\partial/\partial T$ have zero Lie brackets almost everywhere, in virtue of the Einstein equations. Our construction of a holonomic, explicitly stationary basis is thus complete. To write down the components of the metric in this basis, one only needs to evaluate

$$g\left(\frac{\partial}{\partial T}\otimes\frac{\partial}{\partial T}\right)=\dot{a}_{\mathrm{b}}^2-a'^2=\frac{\mathcal{C}}{a^2}-\frac{\Lambda}{6}a^2\tag{8.2.23}$$

$$g\left(\frac{\partial}{\partial r}\otimes\frac{\partial}{\partial r}\right)=g^*(\mathrm{d}r\otimes\mathrm{d}r)^{-1}=\left(\frac{\Lambda}{6}a^2-\frac{\mathcal{C}}{a^2}\right)^{-1}\tag{8.2.24}$$

There is no $r$, $T$ cross term, because $g(\mathrm{d}r,\partial/\partial T)=0$, so the simple diagonal inversion in calculating $g(\partial/\partial r,\partial/\partial r)$ can be justified. Therefore, the metric satisfies

$$g=-\left(\frac{r^2}{\ell^2}-\frac{\mathcal{C}}{r^2}\right)\mathrm{d}T\otimes\mathrm{d}T+\left(\frac{r^2}{\ell^2}-\frac{\mathcal{C}}{r^2}\right)^{-1}\mathrm{d}r\otimes\mathrm{d}r+r^2\gamma_{ij}\mathrm{d}x^i\otimes\mathrm{d}x^j,\tag{8.2.25}$$

where $\ell$ is the AdS curvature scale, defined by $\ell^{-2} = \Lambda/6$. This is the metric of an AdS black hole in five dimensions, with mass $\mathcal{C}$. The crucial point of important in the foregoing derivation is that the coordinate change $r = a$ is only valid provided that $a$ is single valued. As we now describe, there is a critical point at a position $y = y_{\rm h}$ in the bulk at which $a$ has a turning point. The Gaussian normal description makes sense only provided $y < y_{\rm h}$.

The classical solution for $a$ is (Binetruy et al., 2000a,b)

$$\left(\frac{a}{a_{\rm b}}\right)^2 = \sqrt{\left(\frac{2w^2}{(1-w^2)^2} + \frac{\mathcal{C}'}{a_{\rm b}^4}\right) + \frac{2\mathcal{C}'}{a_{\rm b}^4}} \cosh 2\ell^{-1}(y_{\rm h} - y) - \left(\frac{2w^2}{(1-w^2)^2} + \frac{\mathcal{C}'}{a_{\rm b}^4}\right). \quad (8.2.26)$$

The parameter $\mathcal{C}'$ is a modified black hole mass defined by $\mathcal{C}' = \mathcal{C}\ell^2/2$, and for convenience of expression we are using the Hawkins–Lidsey variable $w$ (Hawkins and Lidsey, 2001) to encode the matter theory, where $w$ is defined by

$$\rho = \frac{2\lambda w^2}{1-w^2}. \quad (8.2.27)$$

This expression for $a/a_{\rm b}$ has no zeroes, provided $\mathcal{C}' > 0$, but posesses a turning point at $y = y_{\rm h}$, defined by

$$\tanh 2\ell^{-1} y_{\rm h} = \frac{1+w^2}{1-w^2}\left(\frac{1+w^4}{(1-w^2)^2} + \frac{\mathcal{C}'}{a_{\rm b}^4}\right)^{-1}. \quad (8.2.28)$$

For small $\mathcal{C}'$, $a/a_{\rm b}$ behaves near $y = y_{\rm h}$ like

$$\left(\frac{a}{a_{\rm b}}\right)^2 \overset{y\to y_{\rm h}^-}{\longrightarrow} \frac{(1-w^2)^2}{2w^2}\frac{\mathcal{C}'}{a_{\rm b}^4}\left(1 - \frac{\mathcal{C}'}{a_{\rm b}^4}\frac{(1-w^2)^2}{2w^2} + O(\mathcal{C}'^2)\right). \quad (8.2.29)$$

(These are the small black holes familiar from the study of AdS/CFT.) This condition provide appropriate constraints on the class of fields $a$ which are allowed to enter the quantum model. Suppose one agrees to deal with a particular theory containing a bulk black hole of mass $\mathcal{C}$. In a state where the density of matter on the brane is described by $w$, which in the conformal scalar matter model will be given by an energy eigenvalue $\hbar(m + 1/2)$, one should include functions defined on $(-y_{\rm h}, y_{\rm h})$, where $y_{\rm h}$ satisfies (8.2.28), and with boundary behaviour matching (8.2.29).

In the classical theory, the Gaussian normal coordinates only cover the region up to the first turning point of $a$, specified by (8.2.28). In the following, we assume that for a given matter content, $a$ is specified up $y = y_{\rm h}$. This is an implementation, in the present context, of the principle of black hole complementarity (Susskind, Thorlacius, and Uglum, 1993). In the usual formulation, this principle prescribes that one should formulate a quantum

theory only inside the causal horizon of each observer, possibly with the addition of extra degrees of freedom on the boundary of the causal patch. In the present case the situation is a little different, since the 'Cauchy surface' coinciding with the brane is timelike, and 'Cauchy horizon' at $y = y_{\rm h}$ is not a genuine Cauchy horizon, since it is not the boundary of a causal patch corresponding to initial data given on a Cauchy surface.[3] Nevertheless, this prescription is conventionally used when working with quantum field theory in brane models (Langlois et al., 2000).

It is possible to view this choice as another aspect of the minisuperspace approximation. In the full theory, there would be quantum fluctuations present in the location of the horizon. The fact that the location has been fixed or frozen truncates the quantum theory to more manageable proportions.

One might wonder whether the analysis ought to be undertaken in the global SAdS coordinates (8.2.25). The black hole singularity in these coordinates is not visible from the brane (Mukohyama et al., 2000), although one does then have the problem of keeping track of the event horizon. Moreover, to make contact with analyses which concern gravitational disturbances to the metric (Giudice et al., 2002; Langlois et al., 2000), one should elect local coordinates. It is entirely possible, that the quantization in local coordinates, and the same quantization in global coordinates do not describe the same quantum theory. In that case, one should have to appeal to experiment in order to arbitrate the issue. Unfortunately, since this avenue of experimental investigation seems a long way distant, we are satisfied for the present in working with local coordinates, which at least allows a meaningful comparison with the literature.

**8.2.3. Quantum representation.** We now seek a quantum representation of this theory. In doing so, one should be careful to distinguish between constraints, and field

[3]The surface $y = y_{\rm h}$ has normal form $\mathrm{d}y$ which is everywhere spacelike, even at $y = y_{\rm h}$, and therefore the boundary surface is timelike. Notice that in the absence of time evolution on the brane, and for zero Schwarzschild mass, $n(y)$ and $a(y)$ are zero at $y = y_{\rm h}$, so this is a surface of infinite redshift. On the other hand, adding time evolution or a black hole mass cause $n(t, y) = g_{00}$ to lift from zero at $y = y_{\rm h}$, so the boundary surface is not generically associated with large redshifts. This is important because massive Kaluza–Klein graviton modes, with no Schwarzschild mass and no time evolution on the brane, blow up at $y = y_{\rm h}$. In Chamblin and Gibbons (2000), it was argued that undesirable effects arising from these divergences might be ameliorated owing to the strong redshifting near $y = y_{\rm h}$. In fact the degree of redshifting one can expect depends on how rapidly the brane is evolving.

equations which arise from the variational principle. Given a phase space manifold, field equations pick out preferred trajectories which correspond to the classical evolution. On the other hand, the constraints do not determine trajectories, but instead restrict the allowable combinations of phase space variables which may appear. In the present case, the momentum constraints arising from bulk and boundary variation of $n$ are the relevant constraints, whereas the equations arising from variation of $\pi$ and $a$ are genuine field equations. In such a case one has two options. One can work on the entire phase space, imposing the constraints as operator equations after quantization. Alternatively, one may first reduce the phase space so that the constraints are automatically satisfied, and quantize only afterwards. In general, for non-trivial theories, there does not appear to be any reason to expect the quantum theories following from these distinct procedures to coincide (Carlip, 1998).

In the Randall–Sundrum case, the bulk momentum constraint is complicated and difficult to solve. For this reason, we shall deal with it using the Dirac procedure, after quantizing a relatively unrestricted phase space related to $(\pi, a)$. However, the boundary momentum constraint can be handled much more easily, by restricting phase space to those functions $a$ which satisfy the relevant jump condition. When dealing with gravity coupled to matter on the brane, however, it will be necessary to handle the boundary constraint via the Dirac procedure. This will lead to a brane Wheeler–de Witt equation. In the present pure gravity case, one is dealing with a quantum theory of maps $a : (-y_{\mathrm{h}}, y_{\mathrm{h}}) \to \mathbf{R}$ such that $a$ is even and satisfies the boundary jump condition.

We can now apply the programme outlined in the foregoing remarks. For the purposes of dealing with a quantum cosmological model, it is more appropriate to rotate to Euclidean signature by setting $t = -\mathrm{i}\tau$. We denote the Euclidean action by $I$. The form of $I$ can be considerably tidied up by introducing new variables $v = a^2$ and $N = na$, after which one has

$$I = \int_M \mathrm{d}\tau\, \mathrm{d}y\, \left[\mathrm{i}\pi\dot{v} - N\left(-\frac{2}{\alpha}\pi^2 + \frac{\alpha}{2}\Delta v - \frac{\alpha}{2}\right)\right] - \frac{\alpha}{2}\int_{\partial M} \mathrm{d}\tau N\left(\frac{1}{2}[v']^{+}_{-} + \frac{\lambda}{3}v\right), \quad (8.2.30)$$

where $\Delta$ is the harmonic operator

$$\Delta v = \left(\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2} - \frac{\Lambda}{3}\right)v. \quad (8.2.31)$$

Consider the field equations and constraints arising from this action. The bulk momentum constraint is

$$\mathcal{C}_N = -\frac{2}{\alpha}\pi^2 + \frac{\alpha}{2}\Delta v - \frac{\alpha}{2} = 0. \tag{8.2.32}$$

There is also a boundary momentum constraint, as before,

$$\mathcal{C}_{\partial N} = -\frac{\alpha}{2}\left(\frac{\lambda}{3}v + \frac{1}{2}[v']^+_-\right) = 0. \tag{8.2.33}$$

The bulk variation of $\pi$ gives an expression for the momentum in terms of $\dot{v}$,

$$\pi = -\mathrm{i}\frac{\alpha}{4N}\dot{v}. \tag{8.2.34}$$

There is a $\delta v$ bulk field equation,

$$-\mathrm{i}\dot{\pi} - \frac{\alpha N}{4}\left(\frac{n''}{n} - \frac{2\Lambda}{3}\right) = 0. \tag{8.2.35}$$

(in which $\Delta$ has been replaced by its definition (8.2.31)) and a boundary $\delta v$ constraint,

$$\mathcal{C}_{\partial v} = -\frac{\alpha}{2}\left(N\frac{\lambda}{3} + \frac{1}{2}[N']^+_-\right) = 0. \tag{8.2.36}$$

The constraint $\mathcal{C}_{\partial v}$ just requires that

$$\frac{[a']^+_-}{a} = \frac{[N']^+_-}{N}. \tag{8.2.37}$$

The passage to the quantum representation is now effected in a familiar fashion by representing the Weyl-Heisenberg algebra $[\pi(y), v(y')] = \mathrm{i}\hbar\delta(y - y')$ via operators

$$\pi(y) \mapsto -\mathrm{i}\hbar\frac{\delta}{\delta v(y)} \quad \text{and} \quad v(y) \mapsto v(y) \tag{8.2.38}$$

on a space of functionals of $v(y)$. As noticed above, $v$ should be taken to be defined on a range $(-y_{\mathrm{h}}, y_{\mathrm{h}})$ in agreement with (8.2.28). If the brane is empty and there is no dark radiation, as here, this range reduces to $(-\infty, \infty)$.

**8.2.4. Solution of the quantum constraints.** The bulk momentum constraint is

$$\left(\hbar^2\frac{\delta^2}{\delta v(y)^2} + \frac{\alpha^2}{4}\Delta_y v(y) - \frac{\alpha^2}{4}\right)\Psi[v] = 0. \tag{8.2.39}$$

This is a functional differential equation of Airy form. The term independent of $v$ is absent in a flat model. One supposes an integral solution of Laplace type,

$$\Psi[v] = \int [\mathrm{d}u]\, F[u] \exp \int u(y)v(y)\, \mathrm{d}y, \tag{8.2.40}$$

where the integral is taken over an appropriate class of functions $u(y)$. Since $v$ is specified only on $(-y_{\mathrm{h}}, y_{\mathrm{h}})$ (which may be an infinite range), it suffices that $u$ itself is defined on

$(-y_{\mathrm{h}}, y_{\mathrm{h}})$. In that case, the range of integration is everywhere finite. In the special case that the brane is empty, which is the circumstance under discussion, and there is no bulk Schwarzschild mass, then $v$ is defined on all of $\mathbf{R}$ and satisfies $v \to 0$ as $|y| \to \infty$. In this case, one can conservatively demand that $u$ decays sufficiently fast near infinity that the relevant integrals converge. In addition, since $v$ is even, excepting a discontinuity in derivative at $y = 0$, the functions $u(y)$ can be chosen to be even without loss of generality.

On substituting in the Wheeler–de Witt equation (8.2.39), one obtains

$$0 = \int [\mathrm{d}u] \left[ \hbar^2 u^2(y) + \frac{\alpha^2}{4} \Delta_y v(y) - \frac{\alpha^2}{4} \right] F[u] \exp \int u(y) v(y) \, \mathrm{d}y. \tag{8.2.41}$$

Assuming the validity of a functional integration by parts lemma (or that, given appropriate boundary conditions, the integral of a total variation is zero), this is the same as

$$0 = \int [\mathrm{d}u] \left[ \left( \hbar^2 u^2(y) - \frac{\alpha^2}{4} \right) F[u] - \frac{\alpha^2}{4} \Delta_y \frac{\delta F[u]}{\delta u(y)} \right] \exp \int u(y) v(y) \, \mathrm{d}y. \tag{8.2.42}$$

The term in square brackets $[\cdots]$ should vanish, so $F$ should obey the functional differential equation

$$\Delta_y \frac{\delta F[u]}{\delta u(y)} - \left( \frac{4\hbar^2}{\alpha^2} u^2(y) - 1 \right) F[u] = 0. \tag{8.2.43}$$

To handle the presence of the harmonic operator $\Delta_y$, one defines a set of eigenfunctions $\Omega(m, y)$ by the rule

$$\Delta_y \Omega(m, y) = -\left( \frac{\Lambda}{3} + \frac{m^2}{2} \right) \Omega(m, y). \tag{8.2.44}$$

Bearing in mind that $u(y)$ is even, an appropriate complete set of such eigenfunctions is

$$\Omega(m, y) = \frac{1}{\sqrt{2\pi}} \cos my \tag{8.2.45}$$

where the normalization has been chosen conventionally. If $y_{\mathrm{h}}$ is finite, that the spectrum $\{m\}$ is discrete, whereas if $(-y_{\mathrm{h}}, y_{\mathrm{h}})$ covers $\mathbf{R}$, then the spectrum is continuous. This only involves the exchange of Fourier integrals for Fourier series, so we disregard this subtlety and work entirely in terms of Fourier integrals for convenience. One can express $u(y)$ in terms of a transformed $b(m)$,

$$u(y) = \int \mathrm{d}m \, \Omega(m, y) b(m) \quad \text{and} \quad b(m) = \int \mathrm{d}y \, \Omega(m, y) u(y), \tag{8.2.46}$$

which follows from the reality properties of $\Omega(m, y)$, or, alternatively, evenness of $u(y)$. Functional integration over $u(y)$ is heuristically equivalent to functional integration over

$b(m)$. The variational derivative $\delta/\delta u(y)$ can be expressed in terms of $b(m)$ via

$$\frac{\delta}{\delta u(y)} = \int \mathrm{d}m \, \frac{\delta b(m)}{\delta u(y)} \frac{\delta}{\delta b(m)} = \int \mathrm{d}m \, \Omega(m,y) \frac{\delta}{\delta b(m)}. \tag{8.2.47}$$

The entire functional equation (8.2.43) can be rewritten as an $\Omega$-transform,

$$\Delta_y \int \mathrm{d}m \, \Omega(m,y) \frac{\delta F[b]}{\delta b(m)} - \int \mathrm{d}m \, \Omega(m,y) \left( \frac{4\hbar^2}{\alpha^2} b \star b - \sqrt{2\pi}\delta_{\mathrm{D}}(m) \right) F[b] \tag{8.2.48}$$

where $b \star b$ is the convolution

$$(b \star b)(m) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}z \, b(m-z) b(z). \tag{8.2.49}$$

Therefore, in the $b(m)$ representation,

$$\frac{\delta F[b]}{\delta b(m)} + \left( \frac{\Lambda}{3} + \frac{m^2}{2} \right)^{-1} \left( \frac{4\hbar^2}{\alpha^2 \sqrt{2\pi}} \int \mathrm{d}z \, b(m-z) b(z) - \sqrt{2\pi}\delta_{\mathrm{D}}(m) \right) F[b] = 0. \tag{8.2.50}$$

Writing $F = \exp G$ allows one to deal with the two separate equations

$$\frac{\delta G[b]}{\delta b(m)} = -\frac{4\hbar^2}{\alpha^2 \sqrt{2\pi}} \left( \frac{\Lambda}{3} + \frac{m^2}{2} \right)^{-1} \int \mathrm{d}z \, b(m-z) b(z) \tag{8.2.51}$$

$$\frac{\delta G[b]}{\delta b(m)} = \sqrt{2\pi} \left( \frac{\Lambda}{3} + \frac{m^2}{2} \right)^{-1} \delta_{\mathrm{D}}[m] = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}\omega \, \mathrm{e}^{-\mathrm{i}\omega x} \left( \frac{\Lambda}{3} + \frac{m^2}{2} \right)^{-1}. \tag{8.2.52}$$

Eq. (8.2.52) is fairly simple to solve,

$$G = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}\omega \, \mathrm{d}x \, \mathrm{e}^{-i\omega x} \left( \frac{\Lambda}{3} + \frac{x^2}{2} \right)^{-1} b(x). \tag{8.2.53}$$

This is equivalent, in a distributional sense, to $G \approx (2\pi)^{1/2}(3/\Lambda) b(0)$. We use the symbol $\approx$ to denote weak equivalence, or equivalence up to distributional terms which vanish on integration.

The remaining equation is considerably more complicated to solve, and indeed a general solution will not be possible. By taking Fourier transforms again, it can be shown that (8.2.51) is equivalent to

$$\frac{\delta G[u]}{\delta u(y)} = -\frac{4\hbar^2}{\alpha^2} \int \mathrm{d}z \, \mathrm{j}(z) u^2(y-z), \tag{8.2.54}$$

where $\mathrm{j}(z)$ is the Fourier transform of $(\Lambda/3 + m^2/2)^{-1}$,

$$\left( \frac{\Lambda}{3} + \frac{m^2}{2} \right)^{-1} = \int \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-\mathrm{i}mz} \mathrm{j}(z) \quad \text{or} \quad \mathrm{j}(z) = \sqrt{\frac{3\pi}{\Lambda}} \exp -\sqrt{\frac{2\Lambda}{3}} |z|, \tag{8.2.55}$$

and $u^2(y-z)$ can be written at least formally as $\mathrm{e}^{-z\partial_y} u^2(y)$. Then one is trying to solve

$$\frac{\delta G}{\delta u(y)} = -\frac{4\hbar^2}{\alpha^2} \int \mathrm{d}z \, \mathrm{j}(z) (\cosh z\partial_y) u^2(y). \tag{8.2.56}$$

Unfortunately, the solution of (8.2.56) is far from easy.

Instead of attempting to work with the complicated full quantum theory, described by solutions to (8.2.56), we can instead elect to deal in terms of approximations and estimates. Although information obtained in this way regrettably does not probe the full quantum theory with the same directness as an exact solution of (8.2.56), one can still hope to extract useful details of the gravitational dynamics. In the present case, a useful approximation is suggested by returning to (the Laplace transform of) the Wheeler–de Witt equation (8.2.43). The detailed stucture of this equation is controlled by the coefficient operator $\Delta_y$. If the field $v$ appearing as argument to the Schrödinger functional is a sufficiently mild function of $y$, and $\Lambda \gg 1$, then one may attempt to approximate

$$\Delta_y v(y) \simeq -\frac{\Lambda}{3} v(y). \tag{8.2.57}$$

This is true provided $|v''(y)| \ll 2\Lambda/3$ everywhere. (Unfortunately, this does not include the classical solution, for which $v'' = 2\Lambda/3$ everywhere: see Binetruy et al. (2000b).) In this case, one is trying to solve the very much simpler equation

$$\frac{\delta F[u]}{\delta u(y)} = -\left(\frac{12\hbar^2}{\alpha^2\Lambda} u(y)^2 - \frac{3}{\Lambda}\right) F[u]. \tag{8.2.58}$$

This can be solved directly,

$$F[u] = \exp \int \left(-\frac{4\hbar^2}{\alpha^2\Lambda} u^3(y) + \frac{3}{\Lambda} u(y)\right) \, \mathrm{d}y. \tag{8.2.59}$$

Therefore the full gravitational wavefunctional can be written

$$\Psi[v] \simeq \mathcal{N} \int [\mathrm{d}u] \, \exp \int \left(u(y)v(y) - \frac{4\hbar^2}{\alpha^2\Lambda} u^3(y) + \frac{3}{\Lambda} u(y)\right) \, \mathrm{d}y, \tag{8.2.60}$$

where $\mathcal{N}$ is a normalization constant. The term linear in $u$ is absent in a flat model.

Having recovered an approximate form for the wavefunctional, our next duty is to connect this with known results in terms of four-dimensional or classical physics. The classical result ought to be obtained in the limit $\hbar \to 0$, which sends $\Psi[v] \to \mathcal{N} \int [\mathrm{d}u] \, \exp \int u(v + 3/\Lambda) \, \mathrm{d}y$. A stationary phase approximation then shows that the integral is peaked in the region of $v \simeq -\Lambda/3$, which is just a consistency requirement with (8.2.57). This is the "classical solution" given the approximation which was made, although naturally it does not coincide with the real classical solution, since this wavefunctional is valid in a region of parameter space that does not include the classical case. The proper classical limit can be derived via stationary phase from the exact wavefunctional, as one expects.

It is rather more difficult to extract four-dimensional physics from (8.2.60). This issue will be addressed more fully in the following section, where bulk gravity will be conformally coupled to a scalar field on the brane, and four-dimensional features will more naturally emerge. In the present context, however, one could define a wavefunction of the universe via the rule

$$\psi(v_{\rm b}) \propto \int [{\rm d}v]_{v(y=0)=v_{\rm b}} \Psi[v], \tag{8.2.61}$$

in which five-dimensional bulk features which are unobservable in the four-dimensional effective theory are integrated over. However, it is difficult to make much progress this way, and we will not pursue this approach.

## 8.3. Quantum Randall–Sundrum universe and conformally coupled scalar matter

Although the pure gravitational model is interesting in its own right, there are several elements missing in the present formulation. The bulk analysis should reproduce known features of truncated four-dimensional models (Biswas et al., 2004; Koyama and Soda, 2000; Seahra et al., 2003), which yield a wavefunction for the brane scale factor $a_{\rm b}$ in the Schrödinger representation. In order to uncover these features, it is useful to couple the pure Randall–Sundrum theory to matter living on the brane. A particularly simple, solvable example which has been extensively studied in the four-dimensional case is conformally coupled scalar matter.

**8.3.1. Hamiltonian for RS gravity and conformal scalar.** A conformally invariant scalar field in four dimensions has the (Euclidean) action

$$I_\phi = -\frac{V}{2}\int_{\partial M} {\rm d}\tau \left( \frac{a_{\rm b}^3}{n}\dot{\phi}^2 + n a_{\rm b}\phi^2 + \frac{a^2 \dot{a}_{\rm b}\dot{n}}{n^2}\phi^2 - \frac{a_{\rm b}\dot{a}_{\rm b}^2}{n}\phi^2 - \frac{a_{\rm b}^2 \ddot{a}_{\rm b}}{n}\phi^2 \right) + \text{boundary terms}, \tag{8.3.1}$$

where the boundary terms are chosen to cancel unwanted second derivatives of the metric fields in order to give a properly defined path integral. In the present case, this means that the conformal scalar field action can be reduced to (D'Eath, 1996)

$$I_\phi = -\frac{V}{2}\int_{\partial M} {\rm d}\tau \left[ \frac{a_{\rm b}^3}{n}\left( \dot{\phi} + 2\frac{\dot{a}_{\rm b}}{a_{\rm b}}\phi \right)^2 + n a_{\rm b}\phi^2 \right]. \tag{8.3.2}$$

Following standard methods, we define a new field $\chi = a_{\rm b}\phi$, in terms of which one has

$$I_\phi = -\frac{V}{2}\int_{\partial M} {\rm d}\tau \left( \frac{v_{\rm b}}{N}\dot{\chi}^2 + \frac{N}{v_{\rm b}}\chi^2 \right), \tag{8.3.3}$$

where $v_{\rm b} = v(y = 0)$, and remembering our earlier definition $N = na$. The $\chi^2$ term is absent if the geometry of the universe is flat. It is now easy to rewrite the coupled Randall–Sundrum–scalar system in Hamiltonian form,

$$\begin{aligned} I = &\int_M \mathrm{d}\tau\,\mathrm{d}y \left[\mathrm{i}\pi_v \dot{v} - N\left(-\frac{2}{\alpha}\pi_v^2 + \frac{\alpha}{2}\Delta_y v(y) - \frac{\alpha}{2}\right)\right] - \\ &\int_{\partial M} \mathrm{d}\tau \left[\mathrm{i}\pi_\chi \dot{\chi} - N\left(\frac{3}{\kappa^2\alpha}\frac{\pi_\chi^2}{v} + \frac{\kappa^2\alpha}{12}\frac{\chi^2}{v} + \frac{\alpha\lambda}{6}v + \frac{\alpha}{4}[v']_-^+\right)\right] \end{aligned} \tag{8.3.4}$$

The constraints and field equations arising from this theory consist in a pair of bulk and boundary momentum constraints,

$$\mathcal{C}_N = -\frac{2}{\alpha}\pi_v^2 + \frac{\alpha}{2}\Delta_y v - \frac{\alpha}{2} = 0 \tag{8.3.5}$$

$$\mathcal{C}_{\partial N} = \frac{3}{\kappa^2\alpha}\frac{\pi_\chi^2}{v} + \frac{\kappa^2\alpha}{12}\frac{\chi^2}{v} + \frac{\alpha\lambda}{6}v + \frac{\alpha}{4}[v']_-^+ = 0, \tag{8.3.6}$$

a pair of algebraic equations relating $\pi_v$ and $\pi_\chi$ to their canonical conjugates,

$$\pi_v = -\mathrm{i}\frac{\alpha}{4N}\dot{v} \tag{8.3.7}$$

$$\pi_\chi = \mathrm{i}\frac{\kappa^2\alpha}{6}\frac{v}{N}\dot{\chi}, \tag{8.3.8}$$

a pair of field equations describing the evolution of $v(y)$ and $\chi$ on phase space,

$$-\mathrm{i}\dot{\pi}_\chi - \frac{\kappa^2\alpha}{6}\frac{N}{v}\chi = 0 \tag{8.3.9}$$

$$-\mathrm{i}\dot{\pi}_v - \frac{\alpha}{4}N\left(\frac{N''}{N} - \frac{2\Lambda}{3}\right) = 0, \tag{8.3.10}$$

and finally an auxiliary constraint arising from variation of $v$ on the boundary,

$$\mathcal{C}_{\partial v} = \frac{3N}{\kappa^2\alpha}\frac{\pi_\chi^2}{v} + \frac{\kappa^2\alpha N}{12}\frac{\chi^2}{v^2} - \frac{\alpha\lambda}{6}N - \frac{\alpha N'}{4} = 0. \tag{8.3.11}$$

One can combine the boundary momentum and $\delta v$ constraints to show that

$$\frac{N'}{N} = \frac{v'}{v}. \tag{8.3.12}$$

Although this is the same relation we found in the theory containing pure gravity and no matter, this is not a general property of the Lagrange multiplier $N$; indeed, this simple relationship between $N'$ and $v'$ on $\partial M$ is broken by departures from conformal invariance. Such breakage does not manifest itself here, since we chose the scalar matter to be conformally coupled.

**8.3.2. Quantum representation.** To quantize this system, we choose a coordinate representation and 'solve' the Weyl–Heisenberg algebra by introducing operators

$$\pi_v = -\mathrm{i}\hbar\frac{\delta}{\delta v(y)} \quad \pi_\chi = -\mathrm{i}\hbar\frac{\partial}{\partial\chi} \tag{8.3.13}$$

acting on a space of functionals of $v(y)$ and $\chi$. A generic wavefunctional will have the form

$$\Psi[v;\chi] = \int [\mathrm{d}u]\ \psi(\chi,\partial v,\partial u)F[u]\exp\int u(y)v(y)\ \mathrm{d}y \tag{8.3.14}$$

where by $\psi(\chi,\partial v,\partial u)$ we indicate that the wavefunction in the scalar sector may also depend on boundary data concerning the fields $v$, $u$. The momentum constraints $\mathcal{C}_N$ and $\mathcal{C}_{\partial N}$ are not required to hold separately, but only in the distributional combination $\mathcal{C}_N + \delta_{\mathrm{D}}(y)\mathcal{C}_{\partial N}$. When solving for $\Psi[v;\chi]$ we will seek weak solutions, which satisfy the constraints in a distributional sense. Acting with $\mathcal{C}_N + \delta_{\mathrm{D}}(y)\mathcal{C}_{\partial N}$ on $\Psi$, one obtains

$$\begin{aligned}
&\mathcal{C}_N+\delta_{\mathrm{D}}(y)\mathcal{C}_{\partial N}\Psi[v;\chi] \approx \\
&\int [\mathrm{d}u]\ F[u]\left[\frac{2\hbar^2}{\alpha}\frac{\partial^2\psi}{\partial v^2}\delta_{\mathrm{D}}(y)^2 + \frac{4\hbar^2}{\alpha}\frac{\partial\psi}{\partial v}\delta_{\mathrm{D}}(y)u(y) + \frac{2\hbar^2}{\alpha}\psi u^2(y) + \psi\frac{\alpha}{2}\Delta_y\frac{\delta}{\delta u(y)} - \right. \\
&\left.\frac{\alpha}{2} - \frac{3\hbar^2}{\kappa^2\alpha}\frac{1}{v}\frac{\partial^2\psi}{\partial\chi^2}\delta_{\mathrm{D}}(y) + \frac{\kappa^2\alpha}{12}\frac{\chi^2}{v}\psi\delta_{\mathrm{D}}(y) + \frac{\alpha v}{4}\left(\frac{2\lambda}{3}+\sigma\right)\psi\delta_{\mathrm{D}}(y)\right]\exp\int uv,
\end{aligned} \tag{8.3.15}$$

using the abbreviation $\int uv = \int u(y)v(y)\,\mathrm{d}y$, and where we have set $\sigma = [v']^+_-/v$. In the functional integration by parts, one now acquires a term involving $\partial\psi/\partial u$,

$$\begin{aligned}
\int [\mathrm{d}u]\ \frac{\alpha}{2}\psi F\Delta_y\frac{\delta}{\delta u(y)}\exp\int uv &= -\int [\mathrm{d}u]\ \frac{\alpha}{2}\exp\left(\int uv\right)\Delta_y\frac{\delta}{\delta u(y)}\psi F \\
&= -\int [\mathrm{d}u]\ \frac{\alpha}{2}\exp\left(\int uv\right)\Delta_y\left(\frac{\partial\psi}{\partial u}\delta_{\mathrm{D}}(y)F + \psi\frac{\delta F}{\delta u(y)}\right),
\end{aligned} \tag{8.3.16}$$

Equating distributional parts proportional to the same derivative of $\delta_{\mathrm{D}}(y)$, it can be seen that $\Psi$ satisfies the Wheeler–de Witt equation, on the boundary as well as in the bulk, provided that the following equations hold: a bulk equation, which is the same as the pure gravitational case,

$$\left(\frac{2\hbar^2}{\alpha}u^2(y) - \frac{\alpha}{2}\right)F[u] - \frac{\alpha}{2}\Delta_y\frac{\delta F[u]}{\delta u(y)} = 0; \tag{8.3.17}$$

and a boundary equation which encodes the quantum evolution of the scalar field $\chi$ (or $\phi$), together with its coupling to gravitational degrees of freedom on the boundary,

$$\frac{2\hbar^2}{\alpha}V_\perp\frac{\partial^2\psi}{\partial {v_\mathrm{b}}^2}+\frac{4\hbar^2}{\alpha}\frac{\partial\psi}{\partial v_\mathrm{b}}u_\mathrm{b}-\frac{3\hbar^2}{\kappa^2\alpha}\frac{1}{v_\mathrm{b}}\frac{\partial^2\psi}{\partial\chi^2}+\frac{\kappa^2\alpha}{12}\frac{\chi^2}{v_\mathrm{b}}\psi+\frac{\alpha v_\mathrm{b}}{4}\left(\frac{2\lambda}{3}+\sigma\right)\psi+\frac{\alpha\Lambda}{6}\frac{\partial\psi}{\partial u_\mathrm{b}}=0. \tag{8.3.18}$$

The quantities occuring in the boundary equation are the volume $V_\perp = \delta_\mathrm{D}(0) = 2y_\mathrm{h}$ of the extra transverse dimension; $v_\mathrm{b} = v(y=0)$ and $u_\mathrm{b} = u(y=0)$. In addition, there is a singular distributional piece proportional to $\delta_\mathrm{D}''(y)$ which arises from $\Delta_y$ acting on $\delta_\mathrm{D}(y)$. This piece is weakly zero, and can justifiably be discarded.

The boundary equation separates, as found in earlier analyses (Koyama and Soda, 2000) based on a truncated boundary action. Writing $\psi = U(\partial v, \partial u)C(\chi)$, the $\chi$-dependent piece is subject to a harmonic oscillator equation

$$\frac{\partial^2 C}{\partial\chi^2}-(\Gamma\chi^2-\varepsilon_m)C=0, \tag{8.3.19}$$

where

$$\Gamma=\frac{\kappa^4\alpha^2}{36\hbar^2}\quad\text{and}\quad\varepsilon_m=\frac{\kappa^2\alpha}{3\hbar^2}E_m \tag{8.3.20}$$

and $E_m$ is the energy eigenvalue. The solution is

$$C=\mathcal{N}\exp\left(-\frac{1}{2}\Gamma^{1/2}\chi^2\right)H_m(\Gamma^{1/4}\chi), \tag{8.3.21}$$

where $H_m(z)$ is a Hermite polynomial in $z$ of order $m$, and $E_m$ is subject to the quantization condition

$$E_m=\hbar(m+\frac{1}{2}),\quad m\in\mathbf{Z}^+\cup\{0\}. \tag{8.3.22}$$

This is just the standard quantization condition that arises from the Hermite equation, and follows in exactly the same way as the familiar quantum harmonic oscillator. The form of the remaining gravitational boundary equation can be considerably simplified by a suitable change of variables. Define

$$U(\partial v,\partial u)=\exp\left(-\frac{u_\mathrm{b}v_\mathrm{b}}{V_\perp}\right)\tilde{\mathcal{U}}(\partial v,\partial u), \tag{8.3.23}$$

after which $\tilde{\mathcal{U}}$ should satisfy

$$\frac{2\hbar^2}{\alpha}V_\perp\frac{\partial^2\tilde{\mathcal{U}}}{\partial {v_\mathrm{b}}^2}-\frac{2\hbar^2}{\alpha}\frac{u_\mathrm{b}^2}{V_\perp}\tilde{\mathcal{U}}+\frac{E_m}{v_\mathrm{b}}\tilde{\mathcal{U}}+\frac{\alpha v_\mathrm{b}}{4}\left(\frac{2\lambda}{3}+\sigma\right)\tilde{\mathcal{U}}-\frac{\alpha\Lambda}{6}\frac{v_\mathrm{b}}{V_\perp}\tilde{\mathcal{U}}+\frac{\alpha\Lambda}{6}\frac{\partial\tilde{\mathcal{U}}}{\partial u_\mathrm{b}}=0. \tag{8.3.24}$$

It is clear from the manner in which $u_{\rm b}$ appears in this equation that all dependence on this auxiliary field can be eliminated by a second change of variables,

$$\tilde{\mathcal{U}}(\partial v, \partial u) = \exp\left(\frac{4\hbar^2 u_{\rm b}^3}{\alpha^2 \Lambda V_\perp}\right) \mathcal{U}(v_{\rm b}), \tag{8.3.25}$$

where $\mathcal{U}$ is a reduced wave function. In terms of $\mathcal{U}$, one has

$$\frac{2\hbar^2}{\alpha} V_\perp \frac{\partial^2 \mathcal{U}}{\partial {v_{\rm b}}^2} + \frac{E_m}{v_{\rm b}} \mathcal{U} + \frac{\alpha v_{\rm b}}{4}\left(\frac{2\lambda}{3} + \sigma - \frac{2\Lambda}{3V_\perp}\right)\mathcal{U} = 0. \tag{8.3.26}$$

Loosely speaking, the quantity $\mathcal{U}$ can sensibly be identified with the Wheeler-de Witt wavefunction on the boundary. This is not entirely accurate, since there will typically be a contribution from the bulk term, which couples to $v_{\rm b}$ via the change of variables described above. The comparison with four dimensions undertaken in the following section, however, does show that $\mathcal{U}$ carries most of the $v_{\rm b}$ dependence. We explicitly verify the bulk coupling with a calculable example in Section 8.4, and show that (in that case) it takes the form of a further Airy function.

**8.3.3. The boundary wavefunction.** Assembling the various pieces derived above, the wavefunctional of the conformal Randall–Sundrum–scalar theory takes the form

$$\Psi[v;\chi] = \mathcal{N} \mathrm{e}^{-\frac{1}{2}\Gamma^{1/2}\chi^2} H_m(\Gamma^{1/4}\chi)\mathcal{U}(v_{\rm b}) \int [\mathrm{d}u]\, F[u] \exp\left(\frac{4\hbar^2 u_{\rm b}^3}{\alpha^2\Lambda V_\perp} - \frac{u_{\rm b} v_{\rm b}}{V_\perp}\right) \exp \int uv, \tag{8.3.27}$$

and one may approximate $F[u]$ by the solution (8.2.59), if desired, which is valid only for fields where $v'' \ll 2\Lambda/3$. There are two distinct régimes in which $\mathcal{U}$ may be estimated: in the very early and the very late universe.

In the limit of very small $v_{\rm b}$ (ie., $a_{\rm b} \to 0$, so one is approaching the putative cosmological singularity), the matter energy density diverges and dominates over the 'cosmological constant' contributions which constitute the pure gravitational potential. Thus, one has

$$\frac{\partial^2 \mathcal{U}}{\partial {v_{\rm b}}^2} + \frac{\alpha E_m}{2\hbar^2 V_\perp} \frac{\mathcal{U}}{v_{\rm b}} \simeq 0. \tag{8.3.28}$$

This is entirely consistent with the well-known observation that matter divergences are the cause of principal difficulty when dealing with the cosmological singularity. In this régime, $\mathcal{U}$ satisfies

$$\mathcal{U} \simeq A\sqrt{\frac{\alpha E_m}{2\hbar^2 V_\perp}} v_{\rm b} J_1\left(2\sqrt{\frac{\alpha E_m}{2\hbar^2 V_\perp}} v_{\rm b}\right) + \mathrm{i}B\sqrt{\frac{\alpha E_m}{2\hbar^2 V_\perp}} v_{\rm b} Y_1\left(2\sqrt{\frac{\alpha E_m}{2\hbar^2 V_\perp}} v_{\rm b}\right). \tag{8.3.29}$$

where $A$ and $B$ are arbitrary constants. One can choose $\mathcal{U}$ to obey the de Witt boundary condition $\mathcal{U}(v_{\rm b}=0)=0$ by discarding the divergent solution involving a Neumann function. Similar behaviour was found in an earlier analysis (Biswas et al., 2004). As noticed above, there will typically be contributions from the bulk sector which couple to the boundary scale factor, which are written out in (8.3.23) and (8.3.25). We are assuming that these contributions do not diverge at small $v_{\rm b}$ in such a way as to spoil $\mathcal{U}\to 0$ there. For this reason (8.3.29) does not describe the full $v_{\rm b}$ dependence of the wavefunctional of the universe near $v_{\rm b}=0$, because of the terms coupled to the functional integral.

In the opposite régime where $v_{\rm b}$ is large, corresponding to the late universe, in which the matter energy density has been diluted away, one has

$$\frac{\partial^2\mathcal{U}}{\partial {v_{\rm b}}^2}-\frac{\alpha^2}{8\hbar^2}\frac{v_{\rm b}}{V_\perp}\left(\frac{2\Lambda}{3V_\perp}-\frac{2\lambda}{3}-\sigma\right)\mathcal{U}\simeq 0, \tag{8.3.30}$$

with solution

$$\mathcal{U}\simeq A\,{\rm Ai}\left[\left(\frac{\alpha^2}{8\hbar^2V_\perp}\left[\frac{2\Lambda}{3V_\perp}-\frac{2\lambda}{3}-\sigma\right]\right)^{1/3}v_{\rm b}\right]+B\,{\rm Bi}\left[\left(\frac{\alpha^2}{8\hbar^2V_\perp}\left[\frac{2\Lambda}{3V_\perp}-\frac{2\lambda}{3}-\sigma\right]\right)^{1/3}v_{\rm b}\right]. \tag{8.3.31}$$

This recovers the conventional Airy wavefunction (D'Eath, 1996; Koyama and Soda, 2000; Wiltshire, 1996) familiar from four dimensions and analyses of the braneworld based on a truncated four-dimensional action principle.

At this stage, one would like to compare the boundary Wheeler–de Witt equation (8.3.26) with the four-dimensional result, which was also the Wheeler–de Witt equation for the brane world cosmology found in Koyama and Soda (2000). One does not expect to find all the features of the four-dimensional case: we cannot yet see any coupling from the bulk, but since it will turn out that a large proportion of the familiar features of the cosmological wavefunction can already by seen in $\mathcal{U}$, it is reasonable to suppose that the bulk sector will not alter the behaviour much. We will check this in Section 8.4 below, in the régime where the bulk quantum theory is accessible.

By choosing units in which $\alpha=1/2$, and remembering that $v_{\rm b}=R^2$, one obtains

$$-\hbar^2V_\perp R\frac{\rm d}{{\rm d}R}\left(\frac{1}{R}\frac{{\rm d}\mathcal{U}}{{\rm d}R}\right)-E_m\mathcal{U}-\frac{R^4}{16}\left(\frac{2\lambda}{3}+\sigma-\frac{2\Lambda}{3V_\perp}\right)\mathcal{U}=0. \tag{8.3.32}$$

This is to be compared with the standard result (D'Eath, 1996),

$$-\hbar^2\frac{{\rm d}^2\mathcal{U}}{{\rm d}R^2}+(R^2-H^2R^4-E_m)\mathcal{U}=0. \tag{8.3.33}$$

The $R^2$ term is missing because it corresponds to a sector which properly lives within the functional integral in five dimensions. In four-dimensions there is no analogous bulk sector; one would need to include contributions from the functional integral in order to see the $R^2$ contributions in the brane world. Neglecting operator ordering concerns, which are unimportant semiclassically, and remembering that $V_\perp = 2y_{\rm h}$, gives

$$-\hbar^2 \frac{\mathrm{d}^2\mathcal{U}}{\mathrm{d}R^2} - \frac{E_m}{2\ell \coth^{-1} Y}\mathcal{U} - \frac{R^4}{32\ell \coth^{-1} Y}\left(\frac{2\lambda}{3} + \sigma + \frac{\Lambda}{3\ell \coth^{-1} Y}\right)\mathcal{U} = 0, \qquad (8.3.34)$$

where $Y$ is an abbreviation for the combination

$$Y = \mu\ell\left(1 + \frac{E_m}{\lambda}\right). \qquad (8.3.35)$$

One can expand the inverse hyperbolic functions in terms of $E_m$, which gives

$$\frac{1}{\coth^{-1} Y} = \frac{1}{\tanh^{-1}\frac{1}{\mu\ell}} + \frac{\mu\ell^{-1}\lambda^{-1}E_m}{(\mu^2 - \ell^{-2})\left(\tanh^{-1}(\mu\ell)^{-1}\right)^2} + \cdots \quad \text{if } \Lambda_4 \neq 0, \qquad (8.3.36)$$

or

$$\frac{1}{\coth^{-1} Y} = \frac{2}{\ln 2 - \ln(E_m/\lambda)} - \frac{\lambda^{-1}E_m}{\ln^2(E_m/2\lambda)} + \cdots \quad \text{if } \Lambda_4 = 0. \qquad (8.3.37)$$

One can verify that this tower of corrections in powers of $E_m$ disappears as one sends $\lambda \to \infty$, which is the decoupling limit in which the brane becomes infinitely stiff and no longer responds to influences emanating from the bulk. The Friedmann equation in the presence of brane matter is (Binetruy et al., 2000b)

$$H^2 = \frac{\Lambda_4}{3} + \frac{\kappa_4^2}{3}\rho\left(1 + \frac{\rho}{\lambda}\right) - \frac{1}{R^2} + \frac{\mathcal{C}}{R^4}. \qquad (8.3.38)$$

Although naïvely one sees from this expression that as $\lambda \to \infty$, the quadratic corrections smoothly contract to zero, the situation is in reality somewhat more complicated, since $\Lambda_4$ depends on $\lambda^2$ via (8.2.15) and the effective four-dimensional gravitational coupling $\kappa_4^2$ is defined by $\kappa_4^2 = \mu\kappa_5^2 = \kappa_5^2\lambda/6$. These quantities are fixed during the limiting procedure. Thus, although one properly expects to recover four-dimensional quantities in this limit (with the possible exception of the dark radiation $\mathcal{C}$), it is necessary to be rather careful about how the limit is taken.

This tower of corrections to the contribution of the matter theory does not affect the form of the Wheeler–de Witt equation, which depends only on the relative disposition of the various factors of $R$. However, it certainly does affect the manner in which the matter theory enters the calculation, in a way which continuously retracts to the four-dimensional result as one decouples the brane from the bulk. In this sense, the presence of a tower of

corrections in increasingly high powers of the energy density $E_m$ is exactly analogous to the quadratic corrections which enter the Friedmann equation. The presence of this tower of corrections arises from the presence of a bulk horizon $y_{\rm h}$. This is a central feature of our quantum description, which is framed in local coordinates. By contrast, the treatment in previous work (Biswas et al., 2004; Koyama and Soda, 2000) is framed in terms of global SAdS coordinates. It is natural, as in the four-dimensional treatment, to interpret the modification of gravity as due to interactions with Kaluza–Klein modes in the bulk (Perez-Victoria, 2001). In this context, the presence of a horizon, and the restriction of the quantum description to degrees of freedom which are interior to the horizon, is essential. As we noticed in the introduction, this is the principle of black hole complementarity (Susskind et al., 1993) as applied in the present context.

The other principal distinction between the boundary Wheeler–de Witt equation (8.3.26) and the conventional result is the absence of a term proportional to $R^2$, which encodes the response of the wavefunction to the curvature of the universe. This is absent in Eq. (8.3.26), because in the present context the curvature enters the bulk contribution rather than the boundary theory. Therefore the response of the wavefunction to curvature is encoded in the path integral term, rather than $\mathcal{U}$.

For comparison, we replicate the results of Koyama and Soda (2000) and Biswas et al. (2004).

Biswas et al. (2004) took the Hamiltonian to be, in $D$ dimensions,

$$\hat{H} = M_{\rm P}\left[h^{1/4}\cos\left(\frac{1}{M_{\rm P}}\frac{\partial}{\partial a}\right)h^{1/4} - \frac{T}{D-1}a\right], \tag{8.3.39}$$

where the metric is in five-dimensional Schwarzschild–Anti de Sitter form,

$$\mathrm{d}s^2 = -h(a)\mathrm{d}t^2 + \frac{\mathrm{d}a^2}{h(a)^2} + a^2\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{8.3.40}$$

for $\gamma_{ij}$ a maximally symmetric 3-geometry. The mini-superspace Lagrangian, after the model has been truncated to four dimensions, is

$$L = M_{\rm P}\left[\dot{a}\sinh^{-1}\left(\frac{\dot{a}}{\sqrt{h(a)}}\right) - \sqrt{h(a)+\dot{a}^2} + \frac{T}{D-1}a\right], \tag{8.3.41}$$

where the $T$ appearing here and in the Hamiltonian is the brane tension. These authors then assume that $T = 0$ and find wavefunctionals of the form

$$\psi_j(a) = C_j\phi_j \quad \text{where} \quad \phi_j = h^{-1/4}\exp\left[-(j+\frac{1}{2})\pi a M_{\rm P}\right], \tag{8.3.42}$$

where $j$ is an integer labelling the excitation state of the brane. The fact that $T$ is set to zero makes comparison with our results somewhat problematic.

Koyama and Soda (2000) assume a five-dimensional metric of the form

$$\mathrm{d}s^2 = -N_t^2(r,t)\mathrm{d}t^2 + L^2(r,t)[\mathrm{d}r + N_r(r,t)\mathrm{d}t]^2 + R^2(r,t)\mathrm{d}\Omega_3^2. \tag{8.3.43}$$

The action is taken to be

$$S = -\mu \int \mathrm{d}t R_0^3 \sqrt{N_{t,0}^2 - L_0^2(\dot{r}_0 + N_{r,0})^2} \tag{8.3.44}$$

where a subscript 0 denotes that a quantity is evaluated on the brane. Having obtained the Hamiltonian form, these authors pick a gauge by setting $L = 1$, giving the lapse and shift functions

$$N_r = 0 \quad \text{and} \quad N_t = -\frac{R^2\dot{R}}{\pi_L} \tag{8.3.45}$$

where $\pi_L$ is the momentum canonical to $L$. After dropping some terms from this action and specialising to brane-centred coordinates, the classical equation of motion is obtained to be

$$\left(\frac{\dot{R}_0}{R_0}\right)^2 = -\frac{1}{R_0^2} + H^2 \tag{8.3.46}$$

where $H^2$ is related to the brane cosmological constant. This is exactly equivalent to the four-dimensional Friedmann equation, where there is no matter density. The Wheeler–de Witt equation is

$$\left(-\frac{1}{2}R_0\frac{\partial}{\partial R_0}R_0^{-1}\frac{\partial}{\partial R_0} + V(R_0)\right)\Psi(R_0) = 0, \tag{8.3.47}$$

where $V(R_0)$ obeys

$$V(R_0) = \frac{1}{2}\left(\frac{9\pi^2}{4G_4^2}\right)R_0^2(1 - H^2R_0^2) \tag{8.3.48}$$

in which $G_4$ is the four-dimensional Newton constant. This is the same as the four-dimensional Wheeler–de Witt equation, missing the tower of corrections in inverse powers of $\lambda$ which we found.

## 8.4. The bulk gravitational sector

Having considered the boundary theory in some detail (both the matter theory and the gravitational degrees of freedom), we return to the bulk gravitational sector of the wavefunction (8.3.27). The underlying idea is that since the bulk sector has a description in terms of a functional integral, it can be recast as a particular auxiliary quantum theory. In this representation, the argument $v(y)$ of the wavefunctional can be understood as

a particular time dependent source term, so that the functional integral is in essence a generating functional. There are several options for handling the resulting explicitly time dependent quantum theory. In the one-loop semiclassical approximation, keeping only quadratic fluctuations around the classical solution, it can be managed if desired using well-known Ermakov invariant techniques (Dittrich and Reuter, 1992; Fernández Guasti and Moya-Cessa, 2003) which are appropriate for a time dependent harmonic oscillator. Such techniques also have interesting applications to cosmology (Hawkins and Lidsey, 2002; Rosu, Espinoza, and Reyes, 1999). In this section we employ a more direct approach. It will turn out, in fact, that under the approximations which let us identify the theory with confidence in the $v'' \ll 2\Lambda/3$ limit, the theory is in a strongly coupled, ultra-classical state where the path integral can be evaluated without recourse to detailed calculational techniques.

Suppose now that we restrict attention to fields $v(y)$ such that $v''(y) \ll 2\Lambda/3$, so that the Laplace functional $F[u]$ is given by the approximation (8.2.59). The Schrödinger functional becomes

$$\Psi \propto C(\chi)\mathcal{U}(v_{\mathrm{b}}) \int \mathrm{d}u_{\mathrm{b}}\, \mathrm{d}u_{\mathrm{r}} \int [\mathrm{d}u]_{u_{\mathrm{b}}}^{u_{\mathrm{r}}}\ \exp\left(\frac{4\hbar^2}{\alpha^2 \Lambda V_\perp} u_{\mathrm{b}}^3 - \frac{u_{\mathrm{b}} v_{\mathrm{b}}}{V_\perp}\right) \exp \int \mathrm{d}y\ \left(-\frac{4\hbar^2}{\alpha^2\Lambda} u^3 + uv + \frac{3}{\Lambda} u\right). \tag{8.4.1}$$

The path integral $\int [\mathrm{d}u]_{u_{\mathrm{b}}}^{u_{\mathrm{r}}}$ integrates over all paths between $y = 0$ and $y = y_{\mathrm{h}}$ which satisfy $u(0) = u_{\mathrm{b}}$ and $u(y_{\mathrm{h}}) = u_{\mathrm{r}}$. One can identify the structure of the integrand as giving Airy-like behaviour. There is a bulk Airy functional integral, which couples to an Airy integral over the boundary. This boundary Airy integral has an additional coupling (via $V_\perp$) to the matter theory carried on the brane. The bulk functional integral can be interpreted as the generating functional for a quantum mechanical system with a cubic potential, in the strongly-coupled limit where all kinetic terms have been suppressed. Under this identification, the AdS cosmological constant $\Lambda$ should be identified with the deformation parameter $\hbar$ which controls the correspondence between classical and quantum mechanics. This generating function takes the form

$$Z_3^{(\Lambda)}[v] = \int [\mathrm{d}u]_{u_{\mathrm{b}}}^{u_{\mathrm{r}}}\ \exp \frac{1}{\Lambda} \int \mathrm{d}y\ (-g_3^2 u^3 + u\tilde{v}), \tag{8.4.2}$$

where $\tilde{v}$ is a rescaled source function, $\tilde{v} = \Lambda v + 3$ and $g_3$ is the cubic coupling constant,

$$g_3 = \frac{2\hbar}{\alpha}. \tag{8.4.3}$$

Recall that we are in the Euclidean sector, so one expects exponentials to be damped, not oscillating. This accounts for the factor of i which would be present in the exponential in a Lorentzian theory. Since $g_3$ is proportional to $\hbar$, an expansion in $g_3$ is essentially the $\hbar$-series of the original theory, with the first order term in $g_3$ giving the one-loop correction. However, the most obvious feature of this integral is that it may apparently fail to converge, since cubic $u^3$-type theories have known pathologies. All such concerns essentially arise from the fact that, considering the integral defined by (8.4.2) as a quantum field theory, the energy functional is unbounded below. For this reason, the general $u^3$ theory seems rather physically unsatisfactory. Therefore one might entertain apparently legitimate doubts about the validity of the cubic potential (8.2.59), but in fact these difficulties with the $u^3$ will not cause problems here. The principal mitigating factor is that (8.2.59) (or (8.4.2)) is defined in finite volume: the field $u$ is only defined on $(-y_{\mathrm{h}}, y_{\mathrm{h}})$, which is a finite interval (unless the brane is empty of matter, a degenerate case we shall not consider.) The crucial necessity of a local horizon in keeping the theory finite is evident.

The identification of $\Lambda$ as the deformation parameter, analogous to $\hbar$ in pure quantum mechanics, means that there is a good deal that can be said about the behaviour of $Z_3^{(\Lambda)}$ as a function of $\Lambda$. In particular, the $\Lambda \to 0$ limit, in which the AdS cosmological constant is small, should correspond to the semi-classical region, whereas the opposite régime, in which $\Lambda$ is large, should correspond to the 'ultra-quantum' region.

One can augment the theory described by (8.4.2) away from the strong-coupling régime by adding kinetic terms, or terms involving derivatives of $u$. These terms have been omitted as a result of the approximation in which derivatives of $v$ may be neglected; the generic bulk sector of the wavefunctional will have rather non-trivial kinetic terms present, possibly of all orders in derivatives of $u$ if the general theory (that is, without assumptions about $v''$) can be solved by a derivative expansion. However, for the purposes of studying the strongly coupled theory as it stands, the details of the kinetic terms are not too material. For example, consider a canonical kinetic kernel, corresponding to the minimal generating functional

$$\tilde{Z}_3^{(\Lambda)}[v] = \int [\mathrm{d}u]_{u_{\mathrm{b}}}^{u_{\mathrm{r}}} \exp \frac{1}{\Lambda} \int \mathrm{d}y \left( -\frac{1}{2s^2}\dot{u}^2 - g_3^2 u^3 + u\tilde{v} \right), \tag{8.4.4}$$

where we include a tilde to indicate that this generating functional includes a completion away from strong coupling, and $s$ is a large parameter. The theory of interest with respect to (8.4.2) arises in the limit $s \to \infty$. (See also, for example, a similar scheme in Kabat and Lifschytz (2000).) As a quantum theory with quadratic Hamiltonian and without complicated constraints, (8.4.4) is rather straightforwardly equivalent, via standard arguments, to a canonical theory in the Schrödinger representation. Switching to this representation, the canonical theory described by (8.4.4) has Lagrangian and Hamiltonian

$$L = \frac{\dot{u}^2}{2s^2} + g_3^2 u^3 - u\tilde{v} \quad \text{and} \quad H = \frac{s^2\pi^2}{2} - g_3^2 u^3 + u\tilde{v}, \tag{8.4.5}$$

where, as usual $L = \pi\dot{u} - H$. The Schrödinger representation of this theory is described by wavefunctions obeying the heat equation

$$\left(-\frac{1}{2}\frac{\Lambda^2}{s^2}\frac{\partial^2}{\partial u^2} - g_3^2 u^3 + u\tilde{v}\right)\psi(u;y) = -\mathrm{i}\Lambda\frac{\partial}{\partial y}\psi(u;y), \tag{8.4.6}$$

using the quantization prescription $\pi \mapsto -\mathrm{i}\Lambda\partial/\partial u$ appropriate to the quantum mechanics under discussion. The wavefunction is subject to the boundary condition that the particle sits at $u = u_\mathrm{b}$ at time $y = 0$. Since $\tilde{v}$ is explicitly dependent on $y$, the wavefunction $\psi$ will also carry an explicit $y$-dependence. One can see that the limit $s \to \infty$ is, excepting the right hand side, equivalent to the limit $\Lambda \to 0$, which we know to be the classical limit of this theory. Therefore, as $s \to \infty$ and the kinetic terms disappear, one expects the system to behave in an increasingly classical fashion, as the particle described by the Lagrangian (8.4.5) moves increasingly slowly, and the amplitude for the particle to propagate from $u = u_\mathrm{b}$ at time $y = 0$ to $u = u_\mathrm{r}$ at time $y = y_\mathrm{h}$ becomes increasingly sharply peaked about the classical path. When $s = \infty$ the kinetic terms are not present at all, and the appropriately normalised propagation amplitude is

$$Z_3^{(\Lambda)} = \frac{1}{\sqrt{V_\perp}}\delta_\mathrm{D}(u_\mathrm{r} - u_\mathrm{b}). \tag{8.4.7}$$

This merely expresses the fact that the particle is not moving.

Substituting this expression for the bulk sector back into the wavefunction leaves just an integral over the Airy coupling to the boundary,

$$\text{bulk sector} = \frac{\Psi}{C(\chi)\mathcal{U}(v_\mathrm{b})} \to \int \mathrm{d}u_\mathrm{b}\,\mathrm{d}u_\mathrm{r}\,\exp\left(-\frac{g_3^2}{\Lambda V_\perp}u_\mathrm{b}^3 + \frac{u_\mathrm{b}v_\mathrm{b}}{V_\perp}\right)\frac{\delta_\mathrm{D}(u_\mathrm{r} - u_\mathrm{b})}{\sqrt{V_\perp}}. \tag{8.4.8}$$

Making the substitution

$$z = \left(\frac{3g_3^2}{\Lambda V_\perp}\right)^{1/2} u_\mathrm{b}, \tag{8.4.9}$$

one obtains a simpler integral representation

$$\text{bulk sector} = \left(\frac{\Lambda}{3g_3^2 V_\perp^{1/2}}\right)^{1/3} \int \mathrm{d}z \, \exp\left[-\frac{1}{3}z^3 + v_\mathrm{b} z \left(\frac{\Lambda}{3V_\perp^2 g_3^2}\right)^{1/3}\right]. \tag{8.4.10}$$

There are issues of convergence with the integral if one attempts to interpret the contour of integration as real. In particular, the $z^3$ term diverges at large negative $z$, and the term linear in $z$ diverges at large positive $z$. Instead, one must allow for the possibility that $\int \mathrm{d}z$ should be interpreted as a complex contour integral along a contour which can be deformed to the imaginary $z$-axis, since the integrand is entire in complex $z$-plane except for an essential singularity at $\infty$. The choice corresponds to allowing the function $u$ which enters the functional Laplace transform to itself be complex. (There is nothing suspicious about the addition of this freedom; it means we ought to have begun with a functional Fourier transform rather than a Laplace transform.) With this understanding, one obtains

$$\begin{aligned}\text{bulk sector} &= \left(\frac{\Lambda}{3g_3^2 V_\perp^{1/2}}\right)^{1/3} \int_C \mathrm{d}z \, \exp\left(\frac{1}{3}z^3 + v_\mathrm{b} z \left[\frac{-\Lambda}{3V_\perp^2 g_3^2}\right]^{1/3}\right) \\ &= \left(\frac{\Lambda}{3g_3^2 V_\perp^{1/2}}\right)^{1/3} \mathrm{Ai}\left(v_\mathrm{b}\left[\frac{-\Lambda}{3V_\perp^2 g_3^2}\right]^{1/3}\right),\end{aligned} \tag{8.4.11}$$

where we have disregarded an irrelevant overall normalization factor.

**8.4.1. Probability density and early universe behaviour.** In the early universe limit (Section 8.3.3), the relative probability density for $v_\mathrm{b}$ satisfies, disregarding the matter sector and the overall normalization for convenience,

$$\mathbb{P}(v_\mathrm{b}) \propto v_\mathrm{b} J_1^2\left(2\sqrt{E_m} g_3 \hbar V_\perp v_\mathrm{b}^{1/2}\right) \left|\mathrm{Ai}\left(\left[\frac{-\Lambda}{3g_3^2 V_\perp^2}\right]^{1/3} v_\mathrm{b}\right)\right|^2. \tag{8.4.12}$$

The behaviour of this probability density near $v_\mathrm{b} \simeq 0$ can be understood from inspection of Figure 8.1. It is clear that the bulk coupling does not modify the property $\Psi \to 0$ as the brane scale factor decreases to zero. Although one should be wary of drawing general conclusions, because our approximation $v'' \ll 2\Lambda/3$ is only good in a region of parameter space away from the classical solution, this presumably remains true as one continues the quantum theory describing the bulk dynamic away from strong coupling. At this point, one would like to evaluate the expectation value $\langle v_\mathrm{b} \rangle$ of $v_\mathrm{b}$, in order to compare with the suggestion in Biswas et al. (2004) that quantum effects may stabilize the radius of the brane universe at a size of order the Planck length. Unfortunately, in contrast with the

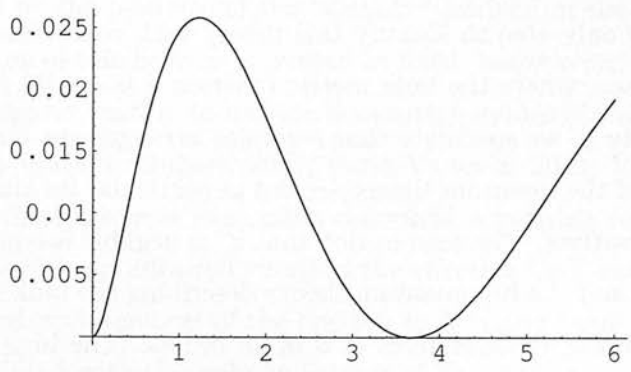



**Figure 8.1.** Relative probability density (vertical axis) near $v_b = 0$ (on the horizontal axis in arbitrary units). This behaviour should be trusted only for $v_b \ll 1$.

analysis carried out in that paper, it is difficult here to properly calculate the expectation value: one must match the early and late approximations for the quantum wavefunctions.

## 8.5. Summary

In this chapter, we have derived the Wheeler–de Witt equation for the brane universe (in a RS-II type scenario with a single domain wall) in local coordinates, and an approximation to the wavefunction describing the cosmology on the domain wall. We find that the wavefunction includes a tower of corrections to the general relativistic result, although these corrections only affect the way the matter theory enters the problem, and not details of the gravitational dynamics. These dynamics are unchanged, leaving the Wheeler–de Witt wavefunction with the same effective functional form as that of four-dimensional gravity, although there is a missing term in our analysis proportional to the square $R^2$ of the radius of the universe which lives in a sector coupled to the bulk in the braneworld. There is an extra sector encoding details of the bulk quantum theory.

We find, in analogy with four dimensions, that one can choose de Witt boundary conditions in which the wavefunction approaches zero as the universe approaches its putative singularity at $R = 0$. In addition, we find that the bulk sector can be described by the strong-coupling limit of a $(0 + 1)$-dimensional quantum field theory. In this limit, the quantum field theory is essentially in its classical régime. The anti-de Sitter cosmological constant plays the role of the quantum deformation parameter $\hbar$ in this theory, allowing one to pick out semiclassical ($\Lambda \to 0$) and ultra-quantum ($\Lambda \to \infty$) regions of the theory.

Although we are only able to identify this theory with confidence on a particular branch of parameter space, where the bulk metric function $v$ is a mild function of the Gaussian normal coordinate $y$, we speculate that a similar arrangement holds for arbitrary $v$, with the parameters of the quantum theory — and in particular its kinetic terms — dependent on $v$ and its derivatives. The assumption that $v''$ is neglible is equivalent to the supression of kinetic terms, and the full quantum theory describing the bulk almost certainly involves kinetic terms containing derivatives of $u$ of all orders. The high-derivative terms in this expansion should be suppressed by powers of the dimensionful constants in the theory, and therefore one might expect that it will be possible to study this theory in the effective "low energy" régime without sending the theory entirely to the strongly coupled limit. This avenue of investigation remains open for future work.

Without solving for the bulk dynamics in detail, we argue that in the strongly-coupled régime the quantum theory describing the bulk wavefunction is essentially in an ultra-classical state where the theory becomes non-propagating. The amplitude corresponding to the generating functional which controls details of the bulk theory can be explicitly evaluated without use of detailed calculational techniques. In this régime, the bulk coupling to the boundary can be explicitly evaluated and contributes another Airy factor to the wavefunction. We show that the de Witt boundary condition can still be maintained, and explicitly exhibit the relative probability density for small values of the scale factor. Unfortunately it is not possible to calculate an expectation value for the scale factor, which would entail knowledge of the probability distribution away from regions where our approximation gives control over it. In fact, the necessity of using this approximation means that only a very small set of observables are calculable in the present framework, so making physical predictions of relevance or interest to observational physics is currently elusive.

The presence of a tower of corrections in coupling the matter theory to gravity is expected in light of the known early-universe modifications of cosmological evolution which are induced by Randall–Sundrum gravity. As a power series in the matter energy density, these corrections become increasingly irrelevant in the late universe where matter density is diluted away by cosmological expansion, but are important in the early universe. Such early time corrections were anticipated in Gubser (2001). However, the general features of the four-dimensional gravitational dynamics are retained. Moreover, finiteness of the

theory is guaranteed by the presence of the "Cauchy" horizon in the bulk. In the present treatment, the location of this horizon is treated as fixed, but one can anticipate extending the minisuperspace approximation to include fluctuating modes of this horizon. In light of the AdS/CFT interpretation (Gubser, 2001; Perez-Victoria, 2001; Verlinde, 2000) of the brane-world, where the transverse dimension described a running renormalization group scale, the horizon cutoff is an infra-red cutoff in the effective CFT and so one expects any divergences associated with removal of the horizon to be fairly tame. On the other hand, the divergence associated with the unboundedness of the cubic potential which describes the bulk sector in the infinite volume limit occurs when the density of matter carried by the brane is zero. This is equivalent to removing the ultra-violet cutoff provided by the brane, which truncates the near-boundary part of AdS. Of course, there is nothing malignant about this procedure, because AdS/CFT guarantees that the conformal field theory is still perfectly well defined. However it is not quite trivial to see how this is reflected in the present formalism. Rather, we presume that an attempt to describe the situation by quantized low energy gravity fails, and one must instead appeal to a higher embedding in string theory for a resolution.

CHAPTER 9

# Winding modes and other exotica

In this chapter, we indicate directions in which braneworld physics can be pushed, and describe some partially completed avenues of investigation. The intention here, unlike the balance of the previous chapters, is not to present finished research or convey firm conclusions (although there are some of these), but instead to complete the task of setting the principal thesis material (Chapters 6–8) in context, by setting the groundwork for progress.

There are some outstanding issues which dominate any list of directions in which to proceed. For one thing, the embedding of the kind of cosmological scenarios which we have been discussing in any fundamental string theory is rather uncertain. There is a great need to find more satisfactory embeddings in order to promote confidence that we are really saying something about the low energy régime of string theory. A rather more elementary version of the same problem is to find more exact solutions of Einstein's equations for the braneworld. Exact solutions are useful here for much the same reasons that they are useful in general (Stephani, Kramer, MacCallum, Hoenselaers, and Herlt, 2003): they are analytically tractable, giving insight into more general models; they provide techniques for matching and approximation; and the offer opportunities to keep methods applicable in more sophisticated cases, like perturbation series, under control. In addition, they are testing grounds for theoretical speculation.

More generally, there are divergent directions in which to proceed. One can concentrate on phenomenology, which is attractive in light of the recent spate of precision cosmological measurements, and bearing in mind the wealth of data which will soon be tapped by the Large Hadron Collider (LHC) at CERN and the Planck satellite coupled to large scale structure and weak lensing observations. Gravitational wave observatories and microwave background polarization experiments may also have opportunities to say something about fundamental physics, and the continuing experimental effort to detect CP violation via B-physics (collaborations such as BaBar and Belle) or pin down the exact

pattern and mechanism by which neutrinos gain mass offers possibilities for new phenomena, new physics, and the necessity of new theoretical explanations which have not really existed since the early days of studying the strong force. This direction is represented in Section 9.4 below, in which we attempt to say something rather general about the gravitational phenomenology of braneworlds by studying Birkhoff's theorem. The principal interest here is the study of braneworld black holes, which is a potentially useful discriminant between models, both in the early and late universe, and is an area in which the difference in dimensionality cannot necessarily be neglected at low energies, as is the case with usual field theory phenomenology.

The technology we use to study Birkhoff's theorem relies on a decomposition of a quite general braneworld into three-dimensional gravity plus a scalar field. Since three-dimensional gravity is well-known to have a Chern–Simons formulation (Carlip, 1998), this prompts speculation about the other principal direction, driven by primarily theoretical considerations. In comparison with the phenomenological approach, in which one attempts to build up to string theory, high energy physics, or quantum gravity effects by parametrizing the results of real-world experiments, the theoretical approach descends in the opposite direction, beginning with a putative quantum gravity, plus compactification scenario, and attempts to derive signatures and characteristics of the ultra-violet theory which should still be visible in the low-energy world. This strand of reasoning is represented here by the next section, in which we attempt to include the winding modes of the string spectrum described in Section 3.3.

## 9.1. Gravitational winding modes

In Chapter 3 is was described how an equivalence between tori of radius $R$ and $R' = \alpha'/R$ exists in string theory, as a consequence of the fact that a string may wrap the circle any number $w \in \mathbf{Z}$ of times. The mass spectrum of string excitations is governed by (3.3.7), which arises in old canonical quantization by imposing the physical state condition $T_{ab} = 0$,

$$m^2 = \frac{n^2}{R^2} + w^2 \frac{R^2}{\alpha'^2}. \tag{9.1.1}$$

Here, $m$ is the excitation mass; $n$ labels the Kaluza–Klein oscillator; and $w$, as noted above, is the winding number. By inspection, it is clear that the spectrum is invariant under the T-duality transformation $n \leftrightarrow w$, $R \leftrightarrow \alpha'/R$ which exchanges momentum and

winding modes and inverts the size of the circle (cf. (3.3.8)). This duality in fact extends beyond the mass spectrum to the entire interacting string theory, and gives a complete equivalence between string theory compactified on a circle of radius $R$ and string theory compactified on a circle of radius $\alpha'/R$. The spectrum of light $w$-excitations which appear as one attempts to contract away the circle ($R \to 0$) will dominate the physics at small $R$ and destroys naïve expectations based on the physics which dominates at large $R$.

As we described in Chapter 5 and Chapter 3, T-duality of string theory is ultimately the reason for the appearance of membranes in the excitation spectrum, which we have used in earlier chapters to discuss cosmological compactifications (Horava and Witten, 1996a,b; Lukas et al., 1999a,b,c). As described above, T-duality exists strictly only for dimensions compactified over tori with all background fields (the gravition $g_{ab}$, the anti-symmetric tensor or Neveu–Schwarz 2-form $B_{ab}$, the dilaton $\phi$ and the vector $A_a$) switched off. In cases where there is a non-trivial background, the fields transform in more complicated ways (summarised in Johnson (2003)). In most cases of cosmological interest, such as the Randall–Sundrum type models (Randall and Sundrum, 1999a,b) with flat branes, and their generalizations with curved branes (Bowcock et al., 2000; Langlois et al., 2000; Mukohyama et al., 2000; Seery and Taylor, 2003), there are non-trivial background fields which one should take into account, so one would not expect the simple T-duality described above between scales of size $R$ and scales of size $\alpha'/R$.[1] Moreover, these models are often phenomenological in nature and do not carry all the information contained in the full interacting string theory, or even a consistent truncation of it. Nonetheless, they are interesting and significant in their own right, and merit serious consideration. However, one would like to include, at least in some measure, important features of the string theory which inspired these models, so there are good reasons to consider what kind of contribution winding modes might make. For example, models exist in which a brane collision is the origin of the high energy density epoch of cosmic evolution usually identified with the Big Bang (Khoury et al., 2002a, 2001; Steinhardt and Turok, 2001) (see also Kanno, Sasaki,

[1]These models usually encode the familiar $\mathbf{S}^1/\mathbf{Z}_2$ orbifold symmetry which comes from the Hořava–Witten model in eleven dimensions, so it is important to ask if winding modes exist on orbifolds. In fact, this is the case: for example, starting with the Type I superstring and compactifying $X^9$ on a circle of radius $R$, one gets a $T_9$-dual theory on the line interval $\mathbf{S}^1/\mathbf{Z}_2$, which is exactly the type of orbifold we are interested in for cosmological purposes. The $\mathbf{Z}_2$ acts in the familiar way by $X^9 \mapsto -X^9$, and the $\mathbf{S}^1$ has radius $R' = \alpha'/R$ (see, eg., Johnson (2003).)

and Soda (2003); Lehners and Stelle (2003)). In these cases, one might be concerned about large numbers of degrees of freedom which become light as the branes approach each other. These degrees of freedom do not appear in field theory models of the collision. For example, winding modes are a feature of a recent attempt to follow details of a brane collision in Turok et al. (2004). This calculation is belived to be relevant to the cyclic and Ekpyrotic brane scenarios.

One approach to finding the spectrum of winding modes might be to attempt to solve the string equations of motion on the braneworld background directly, in which case winding modes would appear naturally, via the analogue of (3.3.1). As is well known, quantum string theory on the 5-dimensional space described by metrics such as (5.1.1) would not make sense in virtue of the Virasoro anomaly, but one would still expect to be able to extract interesting information from the classical string spectrum. But even this limited objective is a daunting prospect. The Polyakov action is (cf. (3.1.1))

$$S_{\rm P} = -\frac{1}{4\pi\alpha'}\int {\rm d}^2\sigma\ (-\gamma)^{1/2}\gamma^{ab}\partial_a X^\mu \partial_b X^\nu g_{\mu\nu}(X) \tag{9.1.2}$$

where $X^\mu(\sigma,\tau)$ describes the embedding of the string worldsheet into spacetime; $\sigma$ and $\tau$ are coordinates on the worldsheet, with indices $a$, $b$, ..., area element ${\rm d}^2\sigma$ and metric $\gamma_{ab}$; and $g_{\mu\nu}$ is the metric on spacetime. For strings governed by (9.1.2), the equations of motion are

$$g_{\mu\nu}(X)\triangle_2 X^\mu + \gamma^{ab}\partial_a X^\mu \partial_b g_{\mu\nu}(X) = 0 \tag{9.1.3}$$

where $\triangle_2$ is the worldsheet Laplacian. Solving this equation is not trivial.

As an alternative which includes some of the right physics, one can consider particle motion on the background described by (5.1.1). Consider some particle worldline $X(\tau)$ parameterized by $\tau$. Then if the free particle Lagrangian is $\mathcal{L}(X)$, one may add a topological term $\mathcal{L}_{\rm WZ}$ which satisfies

$$\int_{\tau_a}^{\tau_b}\mathcal{L}_{\rm WZ} = \frac{\theta}{2y_{\rm h}}\int_{\tau_a}^{\tau_b}{\rm d}\tau\ \frac{{\rm d}X^y}{{\rm d}\tau} \tag{9.1.4}$$

where $\tau_a$ and $\tau_b$ are the endpoints of the motion, $X^y$ is the $y$-component of $X$, and $\theta$ is (at this stage) an arbitrary constant. Since $\mathcal{L}_{\rm WZ}$ is a total derivative, it makes no difference to the classical equations of motion. However, for a particle path which wraps the $y$-dimension a number $w$ times, $\mathcal{L}_{\rm WZ}$ makes a contribution $\theta w$ to the action, and therefore also to the phase of the wavefunction in a semiclassical approximation. This is essentially

the Aharonov–Bohm effect. In the path integral, that gives

$$K_{ab} = \int [\mathrm{d}X] \, \exp\left( \mathrm{i} \int_{\tau_a}^{\tau_b} \mathrm{d}\tau \, \mathcal{L}(X) + \frac{\theta}{2y_\mathrm{h}} \int_{\tau_a}^{\tau_b} \mathrm{d}\tau \, \frac{\mathrm{d}X^y}{\mathrm{d}\tau} \right), \tag{9.1.5}$$

so if we demand that all classical paths with the same endpoints are weighted equally, then $\theta$ must obey an analogue of the Dirac quantization condition, $\theta = 2\pi w$, where $w \in \mathbf{Z}$.

Winding modes have previously been considered in the context of extra-dimensional models. Donini and Rigolin (1999) argue that the relevant string theory is Type I, in which the string scale is not fixed by the four-dimensional Planck scale. They show that for some compactifications, low energy physics may be set by a dominant contributution from low energy string winding modes. In a related context, while this work was in preparation, a preprint by Da Rold (2003) appeared which calculates radiative corrections to vacuum polarization and the quantum effective potential in a similar manner to that described in this section.

Here, we study winding modes of the quantum theory defined by (9.1.5) in the case of both flat and curved branes. In Section 9.2, we solve the Schrödinger equation to find the wavefunctions corresponding to a particular momentum- and winding-number pair $(n, w)$. In particular, we show that the spectrum contains no tachyons, and that the well known mass-gap for curved branes (Gorbunov et al., 2001; Langlois et al., 2000) does not depend on $w$. These are straightforward generalizations of known results in the $w = 0$ sector. In Section 9.2.2 we use these wavefunctions to find the mass spectrum by fitting appropriate boundary conditions at $y = 0$ and at the regulator brane. We give asmptotic formulas for the eigenvalue spacing for large and small separation of the branes. We show that at very large separation the winding modes become heavy and hard to excite, so as $R \to \infty$ they lift out of the spectrum and decouple from the theory, leaving only light momentum modes. However for curved branes at small separation, although the momentum modes become heavy (as expected), the winding modes remain discrete. This implies that there will not be large numbers of light modes appearing in the spectrum. In Section 9.2.3 we use this spectral information to construct a 1-loop effective action using the zeta function technique.

## 9.2. Winding mode wavefunctions

A characteristic feature of brane world cosmologies is that some or all matter (open string modes) is confined to the brane, whereas gravity (closed string modes) propagates freely over the full spacetime. There may be winding modes in both sectors. It is clear that a closed string wrapped along a compact dimension is topologically distinct from a unwrapped closed string. This is the sector relevant for gravity in the bulk. Matter on the brane may also feel winding modes. An open string wrapped around a compact dimension is not topologically distinct from an unwrapped string, because the string may contract and continuously unwrap the cycle. However, the presence of a D3 brane corresponding to our universe modifies this behaviour. If the open string is constrained by a Neumann boundary condition to end on the D3 brane, then an open string wrapped on the compact dimension is topologically distinguishable from an unwrapped string, and winding modes are reintroduced into the spectrum. In this chapter, we concentrate only on winding modes of gravitational excitations, but there is no reason of principle why winding modes of Standard Model particles corresponding to open strings fixed to the brane should not also be important, and indeed there is some evidence that this could be the case (Donini and Rigolin, 1999).

A free point-particle moving in the metric (5.1.1) would have Lagrangian one-form $ds$. However, if one wants to model gravitational excitations, then one should begin with the graviton field equation $\Box\Psi = 0$.[2] To produce a particle action, one interprets this as the Schrödinger equation corresponding to some Hamiltonian $\mathcal{H}$ and Lagrangian $\mathcal{L}$. This Lagrangian is

$$\mathcal{L} = \frac{n^2}{4}\dot{t}^2 + \frac{\mathrm{i}\omega}{2}\dot{t} - \frac{\omega^2}{4n^2} - \frac{1}{4}\dot{y}^2 + \frac{\mathrm{i}\sigma}{2}\dot{y} + \frac{\sigma^2}{4} + \frac{a^2}{4}\dot{\mathbf{x}}^2, \tag{9.2.1}$$

where $\omega$ and $\sigma$ are defined by

$$\omega = 3\frac{\dot{a}}{a} - \frac{\dot{n}}{n} \quad \text{and} \quad \sigma = 3\frac{a'}{a} + \frac{n'}{n}. \tag{9.2.2}$$

[2]This is important. On the one hand there is nothing unexpected about this result, which merely says that gravitons moving on a Randall–Sundrum background with the transverse dimension set to be topologically $\mathbf{S}^1/\mathbf{Z}_2$ do not behave like simple particles. This is correct, because gravitational fluctuations should take into acount the full details of the metric perturbation, rather than the simple requirement that they move on geodesics in the background spacetime.

This Lagrangian is obtained by beginning with the graviton field equation $\Box\Psi = 0$ and interpreting all occurences of $\partial/\partial t$ and $\partial/\partial y$ as momentum operators, whereas all occurences of $t$ and $y$ are position operators. The field equation can then be converted to a Hamiltonian and thence to a Lagrangian as described above. This procedure has not previously appeared in the literature. (It is only an *approximation* to try and pick out the salient features of the winding modes.) We have written the embedding coordinates as $X^\mu = (t, \mathbf{x}, y)$. The appearance of an imaginary component may appear disturbing, but this simply reflects the exchange of energy between the gravitational field, and particles with motion in the $t$- or $y$-directions. In the $t$-direction, this exchange can be interpreted as work done against the ambient expansion of the universe; in the $y$-direction it is an effective renormalization group running of the energy scale (Section 5.4; Gubser (2001); Verlinde (2000))

It is now possible to add the topological term $\mathcal{L}_{\mathrm{WZ}}$. This gives a modified Schrödinger equation

$$\left(-\frac{1}{n^2}\frac{\partial^2}{\partial t^2} - \frac{\omega}{n^2}\frac{\partial}{\partial t} + \frac{\partial^2}{\partial y^2} + \sigma\frac{\partial}{\partial y} + \frac{\theta}{y_{\mathrm{h}}}\frac{\partial}{\partial y} + \frac{\triangle}{a^2} + \frac{\theta\sigma}{2} + \frac{\theta^2}{4y_{\mathrm{h}}^2}\right)\Psi = 0. \tag{9.2.3}$$

This equation is obtained by beginning with topologically modified Lagrangian and reversing the process that led from the effective Schrödinger equation $\Box\Psi = 0$ to the effective Lagrangian, ie., replacing all momentum operators with $\partial/\partial t$ and $\partial/\partial y$ (up to factors of i and $-1$) as appropriate. The coefficient $\theta$ is quantized as $\theta = 2\pi w$, $w \in \mathbf{Z}$ as above. As a result, this Schrödinger equation will support many more states than the unmodified, topologically trivial Schrödinger equation. There will be the momentum states which are found by separation, as before, and there will be topological states coming from the choice of $\theta$. Thus (9.2.3) will have degrees of freedom parametrized by two quantum numbers $(m, \theta)$, rather than just $m$ as in the topologically trivial case.

There are two régimes of interest, where this field equation can be solved exactly, corresponding to a flat Minkowski brane, and a brane carrying a de Sitter cosmology with constant Hubble parameter $H$, which we consider in turn below.

We first consider the case of a Minkowski brane. The metric functions $n$ and $a$ are equal and are given by the exponential warp factor (Randall and Sundrum, 1999b)

$$n = a = \exp\left(-2\ell^{-1}y\right) \quad \text{where } y > 0. \tag{9.2.4}$$

This metric is explicitly static, so $\omega = 0$. Since there is a Killing symmetry in each of the three spacelike directions tangential to the brane, there is no obstacle to Fourier transforming along these directions. This corresponds to the replacement $\triangle \mapsto -k^2$ for a Fourier mode of wavenumber $k$. Because the metric is static, we can also Fourier transform in time, but this is a convenience particular to the case of a flat brane and will disappear in more general models. The wavefunctions then take the form

$$\Psi = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \sum_m \mathcal{E}_{m,\theta}(y) \exp\left[\mathrm{i}(Et - \mathbf{k}\cdot\mathbf{x})\right]. \tag{9.2.5}$$

subject to the usual relativistic dispersion relation

$$E^2 - k^2 = m^2. \tag{9.2.6}$$

The $m$-mode wavefunction $\mathcal{E}_{m,\theta}$ satisfies

$$\mathcal{E}''_{m,\theta} - \left(\frac{4}{\ell} - \frac{\theta}{y_\mathrm{r}}\right)\mathcal{E}'_{m,\theta} + \left(\frac{\theta^2}{4y_\mathrm{r}^2} - \frac{2\theta}{\ell y_\mathrm{r}} - m^2 \mathrm{e}^{2\ell^{-1}y}\right)\mathcal{E}_{m,\theta}, \tag{9.2.7}$$

with $m$ chosen to fit appropriate boundary conditions at the brane and regulator, to be discussed below. We make the change of dependent variable

$$\mathcal{E}_{m,\theta} = \exp\left(-\frac{\theta}{2}\frac{y}{y_\mathrm{r}}\right)\mathrm{e}^{2\ell^{-1}y}\phi_m, \tag{9.2.8}$$

and it is convenient to change variable to a conformal bulk coordinate $z$ defined by $\mathrm{d}y = n\,\mathrm{d}z$. In these coordinates, the brane lies at $z = z_\mathrm{b} > 0$, the regulator brane is at $z = z_\mathrm{r}$ and $y = y_\mathrm{h}$ coincides with $z = \infty$. The presence of the regulator brane means that the periodicity which appears in the topological Lagrangian $\mathcal{L}_\mathrm{WZ}$ should no longer be $y_\mathrm{h}$ but $y_\mathrm{r}$. In terms of the $z$ coordinate, this is most conveniently written as

$$y_\mathrm{r} = \int_0^{y_\mathrm{r}} \mathrm{d}y = \int_{z_\mathrm{b}}^{z_\mathrm{r}} n(z)\,\mathrm{d}z = R, \tag{9.2.9}$$

where $R$ can be interpreted as the radius of the extra dimension as measured in the $z$-frame. We adopt this convention for the rest of this chapter.

These changes send the $\mathcal{E}_m$ equation to

$$\frac{\partial^2\phi_m}{\partial z^2} + \frac{1}{z}\frac{\partial\phi_m}{\partial z} + \left(m^2 - \frac{4}{z^2}\right)\phi_m = 0. \tag{9.2.10}$$

This coincides with the field equation in the $w = 0$ sector. We conclude that wavefunctions with higher topological index can be obtained from the $w = 0$ wavefunctions via the

transformation (9.2.8). The $\mathcal{E}_m$ have the explicit form, in the $z$-frame,

$$\mathcal{E}_{m,w} = \exp\left(-\frac{\pi w \ell}{y_{\rm r}} \ln z\right) \mathrm{e}^{2\ell^{-1}y} \left[AJ_2(mz) + BY_2(mz)\right] \tag{9.2.11}$$

where we have set $\theta = 2\pi w$ for $w \in \mathbf{Z}$, and $J_2$ and $Y_2$ are respectively the Bessel and Neumann functions of order 2.

Now consider the case where the brane carries a de Sitter cosmology, with Hubble parameter $H$ (Gorbunov et al., 2001; Langlois et al., 2000). The results will be rather similar to the Minkowski case. In the de Sitter model $n = \mathcal{A}$ and $a = \mathcal{A}a_{\rm b}$, where

$$\mathcal{A} = H\ell \sinh \ell^{-1}(y_{\rm h} - y) \quad \text{and} \quad a_{\rm b} = \mathrm{e}^{Ht}. \tag{9.2.12}$$

There remains a Killing symmetry in each of the brane spacelike directions but time evolution is now non-trivial. The Schrödinger equation (9.2.3) factorizes, with solutions of the form $\Psi = \varphi_m(t;k)\mathcal{E}_{m,\theta}(y)$, where

$$\ddot{\varphi}_m + 3H\dot{\varphi}_m + \left(\frac{k^2}{a_{\rm b}^2} + m^2\right)\varphi_m = 0 \tag{9.2.13}$$

and

$$\mathcal{A}^2\mathcal{E}''_{m,\theta} + \left(4\mathcal{A}\mathcal{A}' + \mathcal{A}^2\frac{\theta}{y_{\rm h}}\right)\mathcal{E}'_{m,\theta} + \left(2\theta\mathcal{A}\mathcal{A}' + \frac{\theta^2}{4y_{\rm h}^2}\mathcal{A}^2 + m^2\right)\mathcal{E}_{m,\theta} = 0, \tag{9.2.14}$$

where overdots denote a derivative with respect to $t$, and primes denote a derivative with respect to the bulk coordinate $y$. This is a straightforward generalization of the technique introduced by Langlois et al. (2000).

Following the pattern established for flat branes, we make the change of dependent variable

$$\mathcal{E}_{m,\theta} = \frac{U_\theta^{1/2}}{\mathcal{A}^2}\phi_m, \tag{9.2.15}$$

where $U$ has the form

$$U_\theta = \exp\left(\theta\frac{y}{y_{\rm h}}\right). \tag{9.2.16}$$

The quantity $\phi_m$ satisfies a simplified differential equation which coincides with the equation (see Langlois et al. (2000)) for $\mathcal{E}_{m,0}$ in the $w = 0$ sector

$$\phi'' - \left(2\frac{\mathcal{A}''}{\mathcal{A}} + 2\frac{\mathcal{A}'}{\mathcal{A}}\frac{\mathcal{A}'}{\mathcal{A}}\right)\phi = -\frac{m^2}{\mathcal{A}^2}\phi. \tag{9.2.17}$$

It now immediately follows that the winding sector wavefunctions with $\theta \neq 0$ are given in terms of the $\theta = 0$ sector by

$$\mathcal{E}_{m,\theta} = \exp\left(\theta \frac{y}{2y_{\rm h}}\right) \mathcal{E}_{m,0}, \tag{9.2.18}$$

which is the same as for the Minkowski brane.

We change variable to the conformal bulk coordinate $z$. The general solution for $\mathcal{E}_{m,\theta}$ is (Gorbunov et al., 2001)

$$\mathcal{E}_{m,w} = \exp\left(-\frac{\pi w}{R}\int_{z_{\rm b}}^{z} \mathcal{A}(z')\,{\rm d}z'\right)\phi_m(z) \tag{9.2.19}$$

and

$$\begin{aligned}\phi_m = AF\left(\begin{array}{c} -\frac{3}{4}+\frac{i}{2}\kappa,\quad -\frac{3}{4}-\frac{i}{2}\kappa \\ 1/2 \end{array}\middle|\, \cosh^2 Hz\right) + \\ B(\sinh^4 Hz)F\left(\begin{array}{c} \frac{5}{4}+\frac{i}{2}\kappa,\quad \frac{5}{4}-\frac{i}{2}\kappa \\ 3 \end{array}\middle|\, -\sinh^2 Hz\right)\end{aligned} \tag{9.2.20}$$

where $\kappa^2 = m^2/H^2 - 9/4$ is a shifted mass eigenvalue; $A$ and $B$ are constants, which may be zero; $F(a,b;c|z)$ is the Gauss hypergeometric function, sometimes written ${}_2F_1(a,b;c|z)$; and, as above, we have written $\theta = \pi w/R$ where $w \in \mathbf{Z}$, in order for the path integral to weight equivalent classical paths equally as outlined in Section 9.1.

At this stage one can immediately adapt the proof outlined in Frolov and Kofman (2002) that the spectrum contains no tachyons, that is, $m^2 > 0$ for all solutions of (9.2.17) obeying appropriate boundary conditions. Equally, it is easy to see using the argument of Gorbunov et al. (2001); Langlois et al. (2000) that there are no normalizable solutions for $m^2 < 9H^2/4$, except for a possible zero mode at $n = w = 0$.

This is our first result: the spectrum contains no tachyons, and respects the mass gap exhibited by the non-winding $w = 0$ sector. In particular this means that, for example, Kaluza–Klein winding modes of the graviton will not be excited during an inflationary epoch because they are heavy with respect to the ambient de Sitter background.

**9.2.1. Fitting boundary conditions.** None of this depends on the boundary conditions. However, the precise eigenvalue spectrum for both flat and curved branes is fixed by the behaviour one chooses to impose on the wavefunctions, both on the brane at $y = 0$, and at the notional regulator brane at some location $y = y_{\rm r} < y_{\rm h}$ in the bulk. This regulator is

presumed to sit closer to the brane than the horizon $y = y_{\rm h}$. (Of course, for the case of a Minkowski brane, $y_{\rm h} = \infty$.) In this case the spectrum will be discrete. As one decouples the regulator by sending $y_{\rm r} \to y_{\rm h}$ the spectrum of eigenvalues will approach a continuum. Therefore in this model we are dealing with the so-called RS-II scenario or "alternative to compactification" which contains just a single brane, although for the purposes of practical computation we will always keep the regulator brane present.

The boundary condition at $y = 0$ is that the anisotropic stress induced on the brane should vanish (Langlois et al., 2000), or $\mathcal{E}'_{m,\theta} = 0$, where a prime denotes a $y$ derivative. These are Neumann boundary conditions. By analogy we take the same boundary condition at the regulator brane $y = y_{\rm r}$, although in principle one could consider the Dirichlet problem or mixed boundary conditions. In any case, the boundary condition at $y = y_{\rm r}$ should become irrelevant as $y_{\rm r} \to y_{\rm h}$.

We first deal with the Minkowski case. In terms of the $\phi_m$, the boundary condition $\mathcal{E}'_{m,\theta} = 0$ (where $' = \partial/\partial y$) becomes

$$\frac{\phi'_m}{\phi_m} = \left( \frac{\pi w \ell}{R} - \frac{2}{\ell} \right) \phi_m, \tag{9.2.21}$$

which when applied at $y = 0$ and $y = y_{\rm r}$ gives a quantization condition on the allowed values of $m$ in each topological sector.

$$\tilde{P}_w(m) = y_2^w m \ell j_2^w(ma/\ell) - y_2^w(m\ell/a) j_2^w(m\ell) = 0 \tag{9.2.22}$$

where

$$y_\mu^w(z) = \left( 2 - \frac{\pi w \ell}{R} \right) Y_2(z) + z Y_2'(z) \tag{9.2.23}$$

$$j_\mu^w(z) = \left( 2 - \frac{\pi w \ell}{R} \right) J_2(z) + z J_2'(z). \tag{9.2.24}$$

These definitions coincide with Flachi and Toms (2001). We have written $\tilde{P}_w(m)$ with a tilde because later we shall wish to rotate to imaginary values of $m$, which we define to be the untilded $P_w$. (9.2.22) is a straightforward modification of the results presented in Flachi and Toms (2001).

Now consider the de Sitter brane. The boundary condition $\mathcal{E}'_{m,\theta} = 0$ (where $' = \partial/\partial y$) becomes

$$\frac{\phi'}{\phi} = \frac{\pi w}{R} \mathcal{A}(z), \tag{9.2.25}$$

where $'$ is now (and for all subsequent equations) defined by $' = \partial/\partial z$. Imposing (9.2.25) at $z = z_{\rm b}$ gives a relation between $A$ and $B$,

$$\frac{A}{B} = \tilde{\Psi}_w = \frac{\frac{\pi w\ell}{R}{\rm s}_{\rm b}^4 F_3^{\rm b} - 4{\rm c}_{\rm b}{\rm s}_{\rm b}^4 F_3^{\rm b} + \left(\frac{25}{8} + \frac{\kappa^2}{2}\right){\rm c}_{\rm b}{\rm s}_{\rm b}^6 F_4^{\rm b}}{\left(\frac{9}{4} + \kappa^2\right){\rm c}_{\rm b}{\rm s}_{\rm b}^2 F_{3/2}^{\rm b} - \frac{\pi w\ell}{R}F_{1/2}^{\rm b}} \tag{9.2.26}$$

where we define

$$F_{3/2}^{\rm b} = F\left(\begin{matrix}\frac{1}{4}+\frac{{\rm i}\kappa}{2}, & \frac{1}{4}-\frac{{\rm i}\kappa}{2} \\ 3/2 \end{matrix}\middle| {\rm c}_{\rm b}^2\right) \qquad F_3^{\rm b}(\rho) = F\left(\begin{matrix}\frac{5}{4}+\frac{{\rm i}\kappa}{2}, & \frac{5}{4}-\frac{{\rm i}\kappa}{2} \\ 3 \end{matrix}\middle| -{\rm s}_{\rm b}^2\right) \tag{9.2.27}$$

$$F_4^{\rm b}(\rho) = F\left(\begin{matrix}\frac{9}{4}+\frac{{\rm i}\kappa}{2}, & \frac{9}{4}-\frac{{\rm i}\kappa}{2} \\ 4 \end{matrix}\middle| -{\rm s}_{\rm b}^2\right) \quad F_{1/2}^{\rm b}(\rho) = F\left(\begin{matrix}-\frac{3}{4}+\frac{{\rm i}\kappa}{2}, & -\frac{3}{4}-\frac{{\rm i}\kappa}{2} \\ 1/2 \end{matrix}\middle| {\rm c}_{\rm b}^2\right) \tag{9.2.28}$$

and we have introduced the highly useful abbreviations ${\rm s}_{\rm b} = \sinh Hz_{\rm b}$, ${\rm c}_{\rm b} = \cosh Hz_{\rm b}$. Imposing (9.2.25) at the regulator brane $z = z_{\rm r}$ gives an eigenvalue equation $\tilde{P}_w(\kappa) = 0$,

$$\tilde{P}_w(\kappa) = {\rm c}_{\rm r}{\rm s}_{\rm r}^2\left[\tilde{\Psi}_w\left(\frac{9}{4}+\kappa^2\right)F_{3/2}^{\rm r} + 4{\rm s}_{\rm r}^2 F_3^{\rm r} - \left(\frac{25}{8}+\frac{\kappa^2}{2}\right){\rm s}_{\rm r}^2 F_4^{\rm r}\right] - \frac{\pi w\ell}{R}\left[\tilde{\Psi}_w F_{1/2}^{\rm r} + {\rm s}_{\rm r}^4 F_3^{\rm r}\right] \tag{9.2.29}$$

where the $F^{\rm r}$ are defined as in Eqs. (9.2.27)–(9.2.28) with quantities evaluated at the brane replaced with quantities evaluated at the regulator, and we have an analogous definition for ${\rm s}_{\rm r}$ and ${\rm c}_{\rm r}$.

**9.2.2. Eigenvalues and the mass spectrum.** The exact form of the eigenvalue equations (9.2.29) and (9.2.22) are rather unwieldy. Although we will return to them when computing 1-loop effective actions in Section 9.2.3, they are of little value for actual computation. A more practical alternative is to seek approximate solutions to (9.2.7) and (9.2.14). In this section, we construct such approximate solutions and use them to find asymptotic estimates for the mass eigenvalues when the mass is large.

We begin from the off-brane wavefunction equation (9.2.7). When solving for the wavefunctions exactly it was convenient to extract a factor of ${\rm e}^{-\theta y/2y_{\rm r}}{\rm e}^{2\ell^{-1}y}$ which reduces the field equation to the Bessel equation. For present purposes, it is more convenient simply to remove the topological factor ${\rm e}^{-\theta y/2y_{\rm r}}$ and change to the conformal bulk coordinate $z$. Accordingly, setting

$$\mathcal{E}_m = \exp\left(-\frac{\pi w\ell}{R}\ln z\right) z^{3/2}\varphi_m \tag{9.2.30}$$

one finds that $\varphi_m$ obeys the equation

$$\varphi_m'' + \left(m^2 - \frac{15}{4z^2}\right)\varphi_m = 0. \tag{9.2.31}$$

At large $m$, or deep in the bulk where $z^{-2} \to 0$, the solution to this equation can be approximated by

$$\varphi_m = \alpha(m)\cos\left[m(z - z_{\rm b}) + \theta(m)\right] \tag{9.2.32}$$

where $\alpha(m)$ is a slowly varying amplitude, and $\theta(m)$ is an $m$-dependent phase which is to be determined. The boundary conditions are

$$\frac{\varphi_m'}{\varphi_m} = \left(\frac{\pi w\ell}{R} - \frac{3}{2}\right)\frac{1}{z}. \tag{9.2.33}$$

Since $\alpha(m)$ is supposed to be slowly varying, imposing this condition at the brane $z = z_{\rm b}$ gives

$$\tan\left[m(z - z_{\rm b}) + \theta(m)\right] = -\frac{\Upsilon(z_{\rm b})}{m}, \quad \text{where} \quad \Upsilon(z_{\rm b}) = \left(\frac{\pi w\ell}{R} - \frac{3}{2}\right)\frac{1}{z_{\rm b}}. \tag{9.2.34}$$

If $m$ is large, sufficiently large that $m \gg \Upsilon(z_{\rm b})$, then

$$\theta(m) \simeq -\frac{\Psi(z_{\rm b})}{m}. \tag{9.2.35}$$

Applying the boundary condition at the regulator $z = z_{\rm r}$ gives

$$\tan\left[m\Delta z + \theta(m)\right] = -\frac{\Upsilon(z_{\rm r})}{m}, \tag{9.2.36}$$

where $\Delta z = z_{\rm r} - z_{\rm b}$. Therefore,

$$m = \frac{n\pi}{\Delta z} - \left(\frac{\pi w\ell}{R} - \frac{3}{2}\right)\frac{z_{\rm b}^{-1} - z_{\rm r}^{-1}}{n\pi} \quad n, w \in \mathbf{Z}, \tag{9.2.37}$$

where $z_{\rm b}^{-1} - z_{\rm r}^{-1} > 0$, since $z_{\rm r} > z_{\rm b}$. This is the analogue of (9.1.1) for the mass of a string excitation with momentum quantum number $n$ and winding number $w$.

For the case of the de Sitter brane, we keep the decomposition (9.2.19) for $\mathcal{E}_{m,w}$ and write $\phi_m = \mathcal{A}^{-3/2}\varphi_m$. This gives (Langlois et al., 2000)

$$\varphi_m'' = -\left(m^2 - \frac{9H^2}{4} - \frac{15}{4}H^2\ell^2\frac{1}{\sinh^2 Hz}\right)\varphi_m \tag{9.2.38}$$

In comparison with the flat brane the situation is slightly different, since if either $m^2$ or $\sinh^2 Hz_{\rm b}$ is very large (ie., we are dealing with asymptotically heavy modes *or* a

brane carrying large tension, which implies $Hz_{\rm b} \gg 1$) then the right hand side is close to $-H^2\kappa^2\varphi_m$, and has approximate solution

$$\varphi_m \simeq \alpha(z)\cos\Big[H\kappa(z-z_{\rm b})+\theta(\kappa)\Big] \tag{9.2.39}$$

where, as before, $\alpha(z)$ is a slowly varying amplitude, and $\theta(\kappa)$ is a mode-dependent phase. The $\varphi$ boundary condition is obtained from (9.2.25) by substituting for $\phi_m$,

$$\frac{\varphi_m'}{\varphi_m} = \frac{\pi w}{R}\mathcal{A} + \frac{3}{2}\frac{\mathcal{A}'}{\mathcal{A}}. \tag{9.2.40}$$

Since $\alpha(z)$ is slowly varying by assumption, $\varphi_m'/\varphi_m$ satisfies

$$\frac{\varphi_m'}{\varphi_m} = H\kappa\tan\Big[H\kappa(z-z_{\rm b})+\theta(\kappa)\Big]. \tag{9.2.41}$$

Applying this at the brane allows one to fix $\theta(\kappa)$,

$$\tan\theta(\kappa) = \frac{\Upsilon(z_{\rm b})}{H\kappa}, \quad \text{where} \quad \Upsilon(z_{\rm b}) = \frac{\pi w}{R}H\ell\frac{1}{\sinh Hz_{\rm b}} - \frac{3H}{2}\coth Hz_{\rm b}. \tag{9.2.42}$$

Under circumstances where this approximation is valid, that is if $m^2$ is large compared with $\sinh^{-2}Hz_{\rm b}$, then $\Upsilon/\kappa$ is small and one can identify $\theta(\kappa) \simeq \Upsilon(z_{\rm b})/H\kappa$. The boundary condition at $z = z_{\rm r}$ gives an equation for $\kappa$,

$$\tan\Big[H\kappa\Delta z + \theta(\kappa)\Big] = \frac{\Upsilon(z_{\rm r})}{H\kappa}, \tag{9.2.43}$$

where $\Delta z = z_{\rm r} - z_{\rm b}$. This means that $\kappa$ explicitly satisfies

$$\kappa \simeq \frac{n\pi}{H\Delta z} - \frac{\Delta\Upsilon}{n\pi} + \mathcal{O}\left(\frac{1}{n^2}\right) \quad n, w \in \mathbf{Z} \tag{9.2.44}$$

where $\Delta\Upsilon = \Upsilon(z_{\rm b}) - \Upsilon(z_{\rm r}) > 0$. If $Hz_{\rm b} \gg 1$ then this is approximately $(w\pi/R)\Delta\mathcal{A}$, where $\Delta\mathcal{A} = \mathcal{A}(z_{\rm b}) - \mathcal{A}(z_{\rm r})$, and so

$$\kappa \simeq \frac{n\pi}{H\Delta z} - \frac{w}{n}\frac{\Delta\mathcal{A}}{R} + \mathcal{O}\left(\frac{1}{n^2}\right) \quad (Hz_{\rm b} \gg 1). \tag{9.2.45}$$

Eq. (9.2.37) and (9.2.45) have the expected similarities to the string case (9.1.1), with some important differences. Firstly, (9.2.37) and (9.2.45) are asymptotic expansions in $n$ which have higher terms, depending on $w$, which could become important as $w \to \infty$. Therefore one should only trust these expansions if $n \gg w$; they are good mostly for small winding numbers. In the case of curved branes, the procedure of fitting approximate eigenfunctions depended only on the assumption that $m^2$ or $\kappa^2$ was large compared with $\sinh^2 Hz_{\rm b}$ so in principle one could consider finding an asymptotic series in $w$ which would be valid for $w \gg n$, but in practice this is complicated because then $\Upsilon$ and $\kappa$ can be of

comparable magnitude. In the case of flat branes the position of the brane is fixed and the expansion is usually good only if $m$ individually is large for any fixed value of the ambient AdS curvature.

Let us deal first with the simpler case of flat branes. Consider the limit where the branes are very far from each other, ie. $z_{\rm r} \to \infty$. In this limit, remembering that $R = \int_{z_{\rm r}}^{z_{\rm b}} {\rm d}z\,(\ell^{-1}z)^{-1} = \ell^{-1}(\ln z_{\rm r} - \ln z_{\rm b})$, the spacing of $w$ eigenmodes at fixed $n$ becomes infinitely large. On the other hand, since $\Delta z \to \infty$, the spacing of $n$ eigenmodes at fixed $w$ approaches zero. This means that Kaluza–Klein states are becoming light and winding states are becoming heavy, and matches what one would naïvely have expected from field theory. In the opposite limit of close approach, $z_{\rm r} \to z_{\rm b}$, one has $\Delta z \to 0$ and the $n$ eigenmodes become infinitely heavy. (The apparent divergence in (9.2.37), or (9.2.45), as $n \to 0$ is not real, because this relation is valid only when $n \gg 1$, or at least $n \gg w$. Therefore this line of reasoning says nothing about the low-$n$ modes directly. However on the basis of (9.2.45) one still expects $n \neq 0$ modes to generically decouple as $z_{\rm r} \to \infty$.) The winding modes behave differently. Their spacing approaches the finite limit $z_{\rm b} = \ell$, so they are not becoming arbitrarily light even as the branes touch.

Now consider curved branes, restricting attention to eigenvalues for which (9.2.45) is a good approximate description. Such eigenvalues have momentum modes which are quantized in units of $\Delta z^{-1}$, so in the large-radius limit where the regulator brane is removed, $z_{\rm r} \to \infty$, these modes become light and approach a continuum. In the limit where $z_{\rm r} \to z_{\rm b}$, they become heavy and are difficult to excite. In the limit of very close approach, momentum modes with higher values of $n$ decouple from low-energy physics and lift out of the spectrum. This is identical with the Minkowski case.

In contrast, the winding modes are quantized in units of $\Delta_w$, where

$$\Delta_w = \frac{\mathcal{A}(z_{\rm b}) - \mathcal{A}(z_{\rm r})}{\int_{z_{\rm b}}^{z_{\rm r}} \mathcal{A}\,{\rm d}z} = \frac{\ell^{-1}\Big[1 - \mathcal{A}(z_{\rm r})\Big]}{\log\coth\frac{Hz_{\rm r}}{2} - \log\coth\frac{Hz_{\rm b}}{2}} \tag{9.2.46}$$

since $\mathcal{A}(z_{\rm b}) = 1$. As $z_{\rm r} \to \infty$, one has

$$\Delta_w \to \frac{1}{\ell\log\tanh\frac{Hz_{\rm b}}{2}}. \tag{9.2.47}$$

This is non-zero provided that $z_{\rm b} > 0$ or, equivalently, $H > 0$. There is still a finite eigenvalue spacing in the limit that the branes recede infinitely far. The presence of tension or curvature on the branes has regulated the behaviour of the eigenvalues in comparison

with the case of flat branes. Now consider the close-approach case $z_r \to z_b$. Here $\Delta_w$ is also finite,

$$\Delta_w \to \frac{\mathcal{A}'(z_b)}{\mathcal{A}(z_b)} = \sqrt{H^2 + \frac{1}{\ell^2}}. \tag{9.2.48}$$

Therefore, the winding modes are still quantized. This is the same as flat branes. As before this says nothing about the low-$n$ modes directly, but one expects that this conclusion also extends to them. In contrast, Eqs. (9.2.47)–(9.2.48) should be a good approximation to the $w$ quantization in each limit for low $w$-modes. Recall that in the case of the string on a circle of radius $R$, this limit would generate a continuum of winding modes. Here, no such continuum appears. The winding modes form a discrete lattice even in the limit of touching branes $z_r \to z_b$.

All the winding modes with $w \neq 0$ have four-dimensional mass greater than $3H/2$ and are therefore heavy with respect to the ambient de Sitter fluctuations. However, one may take the finiteness of (9.2.48) as an indication that there are not large numbers of winding modes which become light as the branes approach each other. To examine this feature more closely, one should try and calculate the attractive force exerted on one brane by the other due to graviton transmission, including winding modes, between them. In the next section, we attempt an estimate of this attraction.

**9.2.3. 1-loop quantum effective action.** As a final application of winding mode physics, we compute the lowest order term in a heat kernel expansion of the 1-loop quantum effective action, which can be accomplished using the spectral information determined in the previous sections. This 1-loop effective action is the Casimir energy induced on the brane by the $(n, w)$ graviton modes in the bulk, and gives a simple estimate of the mutual attraction or repulsion between the branes caused by the presence of the graviton modes.

In this section, we are essentially following the contour integral technique for handling zeta-functions introduced in Bordag, Elizalde, and Kirsten (1996a); Bordag, Geyer, and Kirsten (1996b). This technique was applied to the brane world in Flachi et al. (2001); Flachi and Toms (2001) where similar results to those presented below were obtained for the case of a flat brane. However, our renormalization prescription differs from Flachi et al. (2001); Flachi and Toms (2001), who used the brane tensions as counterterms to cancel divergences in the 1-loop effective action. Here, we shall adopt a different approach (see Bordag, Goldhaber, van Nieuwenhuizen, and Vassilevich (2002)), where we demand that

the Casimir energy vanish as the mass gap in the spectrum is taken to infinity with all other parameters kept fixed.

The 1-loop quantum effective action $\Gamma^{(1)}$ for a quadratic scalar theory $\phi$ satisfies

$$\Gamma^{(1)} = -\frac{1}{2} \ln \det' D \tag{9.2.49}$$

where $\frac{1}{2}\phi D\phi$ is the Lagrangian and $\det'$ is the modified operator determinant, that is, with zero eigenvalues removed. In our case, $D = \sum_w D_w$ where $D_w$ is the corresponding operator in each topological sector. After integrating out the extra dimension, the resulting four-dimensional action will have the form

$$D = \sum_{n,w} \Box + m_{n,w}^2 \tag{9.2.50}$$

The determinant in (9.2.49) satisfies

$$\ln \det' D = \int_0^\infty \frac{\mathrm{d}t}{t} \operatorname{Tr} \mathrm{e}^{-tD} = \int_0^\infty \frac{\mathrm{d}t}{t} \sum_{n,w}{}' \operatorname{Tr} \mathrm{e}^{-t\Box} \mathrm{e}^{-m_{n,w}^2 t}. \tag{9.2.51}$$

The heat kernel $K = \operatorname{Tr} \mathrm{e}^{-t\Box}$ has a well-known universal asymptotic expansion in powers of $t$,

$$\operatorname{Tr} \mathrm{e}^{-t\Box} \simeq \sum_{k \geqslant 0} t^{(k-n)/2} a_k(\Box) \tag{9.2.52}$$

in $n$ dimensions. This expansion exists quite generally, independently of the details of $\Box$ or the background manifold. The coefficients $a_k$ are called the heat kernel coefficients and depend on $\Box$ and the background manifold. One can show that this expansion is equivalent to an expansion in powers of the background curvature. Therefore the first term $a_0$ is curvature-independent and has an interpretation as the Casimir energy. This coefficient has the form

$$a_0 = \frac{1}{(4\pi)^{n/2}} \int_M \mathrm{d}\Omega = \frac{1}{(4\pi)^{n/2}} \operatorname{Vol}(M) \tag{9.2.53}$$

where $M$ is the background manifold with volume measure $\Omega$, and $\operatorname{Vol}(M)$ is the volume of $M$ in this measure. Since the curvature on the brane is typically $R \sim H^2$ higher terms in the heat kernel expansion will also be important, but all such terms depend on the curvature $R$ and hence can be interpreted as modifications to gravity. Here, we focus attention only on the curvature-independent term which shifts the zero-point of the background energy density.

Changing variables in the integration in (9.2.51) then gives the result

$$\Gamma^{(1)} = -\frac{\mathrm{Vol}(M)}{2(4\pi)^{2+s/2}}\Gamma(-2-\tfrac{s}{2})\sum_{n,w} m_{n,w}^{4+s}, \tag{9.2.54}$$

where we have set the dimension of spacetime to be $4+s$. The eigenvalues $m_{n,w}$ accumulate to infinity, so this sum is formally divergent at the physically interesting value $s = 0$. We will evaluate the sum using zeta-function regularization. Here one exploits the auxiliary parameter $s$. When $s$ is sufficiently negative, the sum of eigenvalues converges,

$$\zeta_w(s) = {\sum_{n,w}}' m_{n,w}^{4+s} \quad (s \text{ sufficiently negative}). \tag{9.2.55}$$

One can then define the determinant at $s = 0$ by analytic continuation from the domain where (9.2.55) makes sense. The function $\Gamma^{(1)}(s)$ found by substituting (9.2.55) in (9.2.54) is called the regularized 1-loop effective action,

$$\Gamma^{(1)}(s) = -\frac{\mathrm{Vol}(M)}{2(4\pi)^{2+s/2}}\Gamma(-2-\tfrac{s}{2})\zeta_w(s). \tag{9.2.56}$$

To perform this continuation, one rewrites the sum as a contour integral using the Cauchy theorem,

$$\zeta_w(s) = \frac{1}{2\pi i}\sum_w \oint_{\mathcal{C}} \mathrm{d}m\, m^{4+s}\frac{\partial}{\partial m}\ln \tilde{P}_w(m) \tag{9.2.57}$$

where $\tilde{P}_w(m) = 0$ is the exact eigenvalue equation (9.2.22) or (9.2.29), considered as a function of $m$, depending on whether one wishes to consider flat or curved branes respectively. $\mathcal{C}$ is any contour which encloses the positive zeroes of $\tilde{P}_w(m)$, but excludes the zero mode $m = 0$. For example, one can use an 'inverted' Hankel contour which approaches zero from $\infty$ along the top side of the positive real axis, encircles the origin, and recedes to infinity again along the bottom side of the positive real axis.

First, we work with the Minkowski case. Rotating the contour of integration to the imaginary axis and parametrizing the resulting contour $m = \rho \mathrm{e}^{\pm i\pi/2}$, where $\rho > 0$, we find

$$\Gamma^{(1)}(s) = -\frac{\mathrm{Vol}(M)}{32\pi^2}\frac{1}{\Gamma(3+\frac{s}{2})}\sum_w \int_0^\infty \mathrm{d}\rho\, \rho^{4+s}\frac{\partial}{\partial m}\Bigg[\ln i_2^w(\rho\ell) + \ln k_2^w(\rho a\ell) + \ln\left(1 - \frac{i_2^w(\rho a\ell)k_2^w(\rho\ell)}{i_2^w(\rho\ell)k_2^w(\rho a\ell)}\right)\Bigg]. \tag{9.2.58}$$

This is valid provided $\frac{1}{2} < \mathrm{Re}(s) < 1$ (see Flachi and Toms (2001)). The functions $i$ and $k$ appearing here are defined by

$$i_\mu^w(z) = y_\mu^w(\mathrm{i}z), \quad \text{and} \quad k_\mu^w(z) = j_\mu^w(\mathrm{i}z) \tag{9.2.59}$$

where $y_\mu^w$ and $j_\mu^w$ are the functions of (9.2.23)–(9.2.24). Explicitly, they satisfy

$$i_\mu^w(z) = \left(2 - \frac{\pi w \ell}{R}\right) I_\mu(z) + z I'_\mu(z) \tag{9.2.60}$$

$$k_\mu^w(z) = \left(2 - \frac{\pi w \ell}{R}\right) K_\mu(z) + z K'_\mu(z), \tag{9.2.61}$$

where $K_\mu$ and $I_\mu$ are modified Bessel functions of order $\mu$,

$$I_\mu(z) = Y_\mu(\mathrm{i}z), \quad \text{and} \quad K_\mu(z) = J_\mu(\mathrm{i}z). \tag{9.2.62}$$

The obstacles to convergence of (9.2.58) come from the $\rho \to \infty$ region. Evidently $(\partial/\partial\rho) \ln P_w(\rho)$ should fall off faster than $\rho^{-4}$ as $\rho \to \infty$ if the integral is to converge. To handle this difficulty we adopt the approach of uniform asymptotic expansions expounded in Bordag et al. (1996a,b). This allows us to isolate the various divergences which occur as $\rho \to \infty$. However it has bad behaviour as $\rho \to 0$, so we first isolate this region by introducing some auxiliary scale $Z$ and treating the regions $[0, Z)$ and $[Z, \infty)$ asymmetrically. This is no more than a trivial re-writing of (9.2.58) and $Z$ must cancel out in the final answer.

We adopt the notation of Flachi and Toms (2001). The modified Bessel functions have uniform asymptotic expansions of the form

$$I_\mu(z) = \frac{\mathrm{e}^z}{\sqrt{2\pi z}} \Sigma^I(z) \tag{9.2.63}$$

$$K_\mu(z) = \sqrt{\frac{\pi}{2z}} \mathrm{e}^{-z} \Sigma^K(z), \tag{9.2.64}$$

so $i_\mu^w(z)$ and $j_\mu^w(z)$ are, asymptotically,

$$i_\mu^w(z) = \sqrt{\frac{z}{2\pi}} \mathrm{e}^z \Sigma^I_{\mu,w} \tag{9.2.65}$$

$$k_\mu^w(z) = \sqrt{\frac{\pi z}{2}} \mathrm{e}^{-z} \Sigma^K_{\mu,w}, \tag{9.2.66}$$

where $\Sigma^I_{\mu,w}$ and $\Sigma^K_{\mu,w}$ are given by

$$\Sigma^I_{\mu,w} = \frac{1}{z}\left(2 - \frac{\pi w \ell}{R}\right) \Sigma^I + \Sigma^I - \frac{1}{2z}\Sigma^I + \frac{\mathrm{d}}{\mathrm{d}z}\Sigma^I, \tag{9.2.67}$$

$$\Sigma^K_{\mu,w} = \frac{1}{z}\left(2 - \frac{\pi w \ell}{R}\right) \Sigma^K - \Sigma^K - \frac{1}{2z}\Sigma^K + \frac{\mathrm{d}}{\mathrm{d}z}\Sigma^K. \tag{9.2.68}$$

Adding and subtracting $N$ terms in the asymptotic expansion gives

$$\Gamma^{(1)}(s) = -\frac{\mathrm{Vol}(M)}{32\pi^2} \frac{1}{\Gamma(3 + \frac{s}{2})} \sum_w \left( Z_w(s; a) + \lambda_{\mathrm{b}}^w(s) + \frac{1}{a^{4+s}} \lambda_{\mathrm{r}}^w(s) \right), \tag{9.2.69}$$

where $Z_w(s;a)$ is an $a$-dependent piece,

$$Z(s;a) = \int_0^\infty \mathrm{d}\rho \;\; \rho^{4+s} \frac{\partial}{\partial \rho} \ln\left(1 - \frac{i_2^w(\rho a \ell) k_2^w(\rho \ell)}{i_2^w(\rho \ell) k_2^w(\rho a \ell)}\right) \tag{9.2.70}$$

and the contributions $\lambda_{\mathrm{b}}^w$ and $\lambda_{\mathrm{r}}^w$ are $a$-independent,

$$\begin{aligned}
\lambda_{\mathrm{b}}^w(s) = & \int_0^Z \mathrm{d}\rho \; \rho^{4+s} \frac{\partial}{\partial \rho} \ln i_2^w(\rho\ell) + \\
& \int_Z^\infty \mathrm{d}\rho \; \rho^{4+s} \frac{\partial}{\partial \rho} \left( \ln i_2^w(\rho\ell) - \ln \mathrm{e}^{\rho\ell} \sqrt{\frac{1}{2\pi}} - \sum_{k=0}^{N} \alpha_k (\rho\ell)^{-k} \right) + \\
& \int_Z^\infty \mathrm{d}\rho \; \rho^{4+s} \frac{\partial}{\partial \rho} \left( \ln \mathrm{e}^{\rho\ell} \sqrt{\rho\ell} 2\pi + \sum_{\substack{k=0 \\ k\neq 4}}^{N} \alpha_k (\rho\ell \right)^{-k} - \\
& \frac{k\alpha_4}{s\ell^4} [\rho^s]_Z^\infty ; \\
\lambda_{\mathrm{r}}^w(s) = & \int_0^Z \mathrm{d}\rho \; \rho^{4+s} \frac{\partial}{\partial \rho} \ln k_2^w(\rho\ell) + \\
& \int_Z^\infty \mathrm{d}\rho \; \rho^{4+s} \frac{\partial}{\partial \rho} \left( \ln k_2^w(\rho\ell) - \ln \mathrm{e}^{-\rho\ell} \sqrt{\frac{\pi z}{2}} - \sum_{k=0}^{N} \beta_k (\rho\ell)^{-k} \right) + \\
& \int_Z^\infty \mathrm{d}\rho \; \rho^{4+s} \frac{\partial}{\partial \rho} \left( \ln \mathrm{e}^{-\rho\ell} \sqrt{\frac{\pi z}{2}} + \sum_{\substack{k=0 \\ k\neq 0}} \beta_k (\rho\ell)^{-k} \right) - \\
& \frac{k\beta_4}{s\ell^4} [\rho^s]_Z^\infty
\end{aligned} \tag{9.2.71, 9.2.72}$$

where the coefficients $\alpha_k$, $\beta_k$ are defined by

$$\sum_{k=0}^{\infty} \alpha_k z^{-k} = \ln \Sigma_{2,w}^I(z) \tag{9.2.73}$$

$$\sum_{k=0}^{\infty} \beta_k z^{-k} = \ln \Sigma_{2,w}^K(z). \tag{9.2.74}$$

The (infinite) terms $\lambda_{\mathrm{b}}$ and $\lambda_{\mathrm{r}}$ have the right $a$-dependence to renormalize the brane tensions. The pole terms in $\lambda_{\mathrm{b}}$ and $\lambda_{\mathrm{r}}$ should be discarded. The remaining pieces are analytic for at least $\mathrm{Re}(s) < N-4$. The integrand in $Z_w(s)$ falls off faster than polynomially for $\rho \to \infty$, so this is analytic and one may set $s = 0$ without difficulty. Therefore, the fully renormalized one-loop effective action takes the form

$$S_2 = \lambda_{\mathrm{b}}^{\mathrm{ren}} + \lambda_{\mathrm{r}}^{\mathrm{ren}} - \frac{1}{16\pi^2} \sum_w \int_0^\infty \mathrm{d}\rho \; \rho^3 \ln\left(1 - \frac{i_2^w(\rho a \ell) k_2^w(\rho \ell)}{i_2^w(\rho \ell) k_2^w(\rho a \ell)}\right). \tag{9.2.75}$$

Now consider the de Sitter case. The $\zeta$-function integral (9.2.57) is more conveniently written in terms of the variable $\kappa$ which actually appears in (9.2.29),

$$\zeta_w(s) = \frac{1}{2\pi i} \sum_w \oint_{(0-)}^{\infty} \mathrm{d}\kappa \, \left( H^2\kappa^2 + \frac{9H^2}{4} \right)^{2+s/2} \frac{\partial}{\partial\kappa} \ln \tilde{P}_w(\kappa). \tag{9.2.76}$$

If there is a mode exactly at $m = 3H/2$ then there will be a pole at the origin in the $\kappa$ **C**-plane, so the piece of the contour which encircles the origin contributes $2\pi i \, \mathrm{Res}_{\kappa\to 0}$. In that case, one should carefully remove the zero at $\kappa = 0$ from $P_w(\kappa)$ by multiplying with a suitably chosen function which has no zeroes or poles inside $\mathcal{C}$. The contribution of the pole can be added in by hand, but it is important in what is to follow that the integrand has no singularities on the imaginary axis.

For simplicity, we assume that there is no mode exactly at $m = 3H/2$. Then one can rotate the contour $\oint_{(0-)}^{\infty}$ to the imaginary axis, because there are no singularities in the integrand to prevent this. In particular, $F(a, b; c|z)$ is an analytic function of $a$ and $b$ everywhere except at $a = \infty$ or $b = \infty$, where there is an essential singularity (Erdélyi, 1953). We find, taking account of the branch cut in $z^{s/2}$ at $z = 0$,

$$\Gamma^{(1)}(s) = \frac{\mathrm{Vol}(M)}{32\pi^2} \frac{1}{\Gamma(3 + \frac{s}{2})} \sum_{w=0}^{\infty} \int_{M/H}^{\infty} \mathrm{d}\rho \, \left( H^2\rho^2 - M^2 \right)^{2+s/2} \frac{\partial}{\partial\rho} \ln P_w(\rho), \tag{9.2.77}$$

where we have replaced the mass gap $3H/2$ by a general mass $M$. Our renormalization prescription will be that $\Gamma^{(1)}$ should vanish as $M$ is sent to infinity. In this limit, all modes seen on the brane would be infinitely heavy and one would not expect to see fluctuations. We will verify below that this prescription indeed removes all singularities in (9.2.77). The function $P_w(\rho)$ which controls the pole structure of the integrand is given by the rotation of (9.2.29) to the imaginary axis,

$$P_w(\rho) = \mathrm{c_r s_r^2} \left[ \Psi_w \left( \frac{9}{4} - \rho^2 \right) F^{\mathrm{r}}_{3/2} + 4\mathrm{s_r^2} F^{\mathrm{r}}_3 - \left( \frac{25}{8} - \frac{\rho^2}{2} \right) \mathrm{s_r^2} F^{\mathrm{r}}_4 \right] - \frac{\pi w \ell}{R} \left[ \Psi_w F^{\mathrm{r}}_{1/2} + \mathrm{s_r^4} F^{\mathrm{r}}_3 \right], \tag{9.2.78}$$

and $\Psi_w$ satisfies

$$\Psi_w = \frac{\frac{\pi w \ell}{R} \mathrm{s_b^4} F^{\mathrm{b}}_3 - 4\mathrm{c_b s_b^4} F^{\mathrm{b}}_3 + \left( \frac{25}{8} - \frac{\rho^2}{2} \right) \mathrm{c_b s_b^6} F^{\mathrm{b}}_4}{\left( \frac{9}{4} - \rho^2 \right) \mathrm{c_b s_b^2} F^{\mathrm{b}}_{3/2} - \frac{\pi w \ell}{R} F^{\mathrm{b}}_{1/2}}, \tag{9.2.79}$$

where the $F^{\mathrm{r}}$, $F^{\mathrm{b}}$ are given by the relevant continuations of (9.2.27)–(9.2.28), for example,

$$F^{\mathrm{b}}_{3/2} = F\left( \begin{matrix} \frac{1}{4} - \frac{\varrho}{2}, & \frac{1}{4} + \frac{\varrho}{2} \\ 3/2 \end{matrix} \middle| \, \mathrm{c_b^2} \right) \quad \text{and} \quad F^{\mathrm{r}}_{3/2} = F\left( \begin{matrix} \frac{1}{4} - \frac{\varrho}{2}, & \frac{1}{4} + \frac{\varrho}{2} \\ 3/2 \end{matrix} \middle| \, \mathrm{c_r^2} \right). \tag{9.2.80}$$

In particular, $\Psi_w \to 0$ as $\rho \to \infty$.

The obstacles to convergence of (9.2.77) for general $s$ come from the large-$\rho$ behaviour of the integrand, since the hypergeometric functions $F(a+\lambda, b-\lambda; c|z)$ generically diverge as $\lambda$ approaches infinity on the real axis (see Section 9.3). One can show that the dominant term in (9.2.78) is $F_{1/2}(\rho)$, and that $F_{1/2}(\rho)$ has an asymptotic expansion

$$F_{1/2}(\rho) \sim F^{\infty}_{1/2}(\rho) \sum_{k=0}^{\infty} \alpha_k \rho^{-k} \quad \text{as } \rho \to \infty \tag{9.2.81}$$

for some coefficients $\alpha_k$, and where $F^{\infty}_{1/2}(\rho)$ is defined by (9.3.5). We also introduce associated coeffients $\beta_k$ using the relation

$$\ln \sum_{k=0}^{\infty} \alpha_k \rho^{-k} = \sum_{k=0}^{\infty} \beta_k \rho^{-k}. \tag{9.2.82}$$

This determines the $\beta_k$ in terms of the $\alpha_k$ just by a simple Taylor expansion. By using these definitions and rearranging terms in (9.2.77), one obtains an equivalent form for the regularized effective action,

$$\Gamma^{(1)}(s) = \sum_{w=0}^{\infty} Z_w(s) + A_0(s) + A_1(s) + \sum_{k=0}^{N} W_k(s) \tag{9.2.83}$$

where $Z_w(s)$ is a finite piece, depending on the topological label $w$, defined by

$$Z_w(s) = \frac{\mathrm{Vol}(M)}{32\pi^2} \frac{1}{\Gamma(3+\frac{s}{2})} \int_{M/H}^{\infty} \mathrm{d}\rho \left(H^2\rho^2 - M^2\right)^{2+s/2} \frac{\partial}{\partial\rho} \ln \Bigg[ \mathrm{c_r s_r^2} \rho^{-r} \times$$
$$\left( \left[ \frac{9}{4} - \rho^2 \right] \Psi_w \frac{F_{3/2}}{F_{1/2}} + 4\mathrm{s_r^2} \frac{F_3}{F_{1/2}} - \left[ \frac{25}{8} - \frac{\rho^2}{2} \right] \mathrm{s_r^4} \frac{F_4}{F_{1/2}} \right) - \frac{\pi w \ell}{R} \left( \Psi_w + \mathrm{s_r^2} \frac{F_3}{F_{1/2}} \right) \Bigg]; \tag{9.2.84}$$

the terms $A_1(s)$ and $A_1(s)$ satisfy

$$A_0(s) = \frac{\mathrm{Vol}(M)}{32\pi^2} \frac{1}{\Gamma(3+\frac{s}{2})} \int_{M/H}^{\infty} \mathrm{d}\rho (H^2\rho^2 - M^2)^{2+s/2} \times$$
$$\frac{\partial}{\partial\rho} \left( \ln F_{3/2} - \ln F^{\infty}_{3/2} - \sum_{k=0}^{N} \beta_k \rho^{-k} \right), \tag{9.2.85}$$

$$A_1(s) = \frac{\mathrm{Vol}(M)}{32\pi^2} \frac{1}{\Gamma(3+\frac{s}{2})} \int_{M/H}^{\infty} \mathrm{d}\rho (H^2\rho^2 - M^2)^{2+s/2} \frac{\partial}{\partial\rho} \ln \rho^r F^{\infty}_{3/2}; \tag{9.2.86}$$

and each $W_k(s)$ is the integral of a corresponding $\beta_k$ term,

$$W_k(s) = \frac{\mathrm{Vol}(M)}{32\pi^2} \frac{1}{\Gamma(3+\frac{s}{2})} \int_{M/H}^{\infty} \mathrm{d}\rho (H^2\rho^2 - M^2)^{2+s/2} \frac{\partial}{\partial\rho} \beta_k \rho^{-k}. \tag{9.2.87}$$

Here $N < \infty$ is some integer controlling the domain in which $Z_w(s)$ is analytic, and must be chosen sufficiently large that $Z_w(s)$ is analytic near zero, but is otherwise unimportant; and $r < \infty$ is chosen to make $Z_w(s)$ converge at $s = 0$ (see Section 9.3).

Using simple manipulations, one can show that the $W_k$ integrals can be performed explicitly,

$$W_k(s) = -\frac{\mathrm{Vol}(M)}{32\pi^2} \frac{k\beta_k H^k M^{4-k+s}}{2} \frac{\Gamma(-2 - \frac{s}{2} + \frac{k}{2})}{\Gamma(1 + \frac{k}{2})}. \tag{9.2.88}$$

Therefore, for $k < 4$, the $W_k$ are proportional to a non-negative power of $M$ or $\ln M$ and should be discarded. For $k = 4$ there is a pole at $s = 0$ which appears multiplied by a $M^0$. According to our renormalization prescription, all these terms should be deleted, leaving the $W_k$ for $k \geqslant 5$, which can alternatively be written as

$$W_k(s) = -\frac{\mathrm{Vol}(M)}{32\pi^2} \frac{k\beta_k H^k M^{4-k+s}}{(\frac{k}{2} - 1)(\frac{k}{2} - 2)} \quad (k \geqslant 5). \tag{9.2.89}$$

(In particular, this means that one should take $N > 5$.) This is a finite, analytic function of $s$, so one may set $s = 0$ without prejudice.

Taking into account the asymptotics of $F_{1/2}$, similar reasoning shows that $A_0(s)$ is an analytic function of $s$ in the region $\mathrm{Re}\, s < N - 5$, so there is no obstacle to setting $s = 0$ if $N > 5$. Moreover, this term vanishes as $M \to \infty$, and so must be kept in its entirety. Equally, $A_1(s)$ is proportional to a positive power of $M$ and must be completely discarded. That leaves the function $Z(s)$. This is finite at $s = 0$ and vanishes as $M \to \infty$, so must be kept. This means that the 1-loop Casimir energy can be finally written as, restoring $M = 3H/2$,

$$\Gamma^{(1)} = \left( \sum_{w=0}^{\infty} Z_w(0) + A_0(0) + \sum_{k=5}^{N} W_k(0) \right)_{M=3H/2}. \tag{9.2.90}$$

This is entirely finite; our renormalization prescription has removed all divergences after analytic continuation back to $s = 0$. The terms which have been discarded include all divergences or singularities at $s = 0$. In particular, the form of these divergences or singularities is entirely independent of the topological index $w$. Eq. (9.2.90) is the third principal result of this section.

## 9.3. Asymptotics of the hypergeometric function

In this Appendix we briefly sketch the theory of the asymptotics of the hypergeometric function, and fill in some details concerning the derivation of the Casimir energy, (9.2.90).

In particular, we are concerned with the asymptotics of hypergeometric functions of the form

$$F\left(\left.\begin{matrix} a+\lambda, & b-\lambda \\ & c \end{matrix}\right| z\right) \quad \text{as } \lambda \to \infty, \text{ with } a,\ b,\ c \text{ and } z \text{ fixed.} \tag{9.3.1}$$

This function was first studied by G.N. Watson in 1918 (Watson, 1918), but has also been the subject of recent, more detailed, attention (Jones, 2001; Olde Daalhuis, 2001; Temme, 2001). In particular, Jones (2001) has given a very complete treatment which describes the asymptotics in terms of modified Bessel functions, and gives an associated error bound. In the present case, however, the most convenient form is that due to Watson, who used the method of steepest descent to obtain the result (see also Andrews, Askey, and Roy (2001); Erdélyi (1953); Olver (1974); Wong and Guo (1989))

$$\begin{aligned} F\left(\left.\begin{matrix} a+\lambda, & b-\lambda \\ & c \end{matrix}\right| \frac{1-z}{2}\right) =& \frac{\Gamma(1-b+\lambda)\Gamma(c)}{\Gamma(\frac{1}{2})\Gamma(c-b+\lambda)} 2^{a+b-1}(1-\mathrm{e}^{-\xi})^{1/2-c}(1+\mathrm{e}^{-\xi})^{c-a-b-1/2}\times \\ & \lambda^{-1/2}\left(\mathrm{e}^{(\lambda-b)\xi}+\mathrm{e}^{\mp i\pi(1/2-\gamma)}\mathrm{e}^{-(\lambda+a)\xi}\right)\left[(1+\mathcal{O}(\lambda^{-1})\right] \end{aligned} \tag{9.3.2}$$

where $\mathrm{e}^{\pm\xi} = z \pm \sqrt{z^2-1}$, and one stipulates that

$$(1-\mathrm{e}^{\xi}) = (\mathrm{e}^{\xi}-1)\mathrm{e}^{\mp i\pi} \quad \text{if } \mathrm{Im}(z) \gtrless 0. \tag{9.3.3}$$

The phase $\arg(\lambda)$ satisfies

$$-\frac{\pi}{2} - w_2 + \delta < \arg(\lambda) < \frac{\pi}{2} + w_1 - \delta, \tag{9.3.4}$$

and the various angles occuring here are defined by

$$\begin{aligned} w_2 &= \arctan\frac{\eta}{\zeta}, \quad -w_1 = \arctan\frac{\eta-\pi}{\zeta} \quad \text{if } \eta \geqslant 0 \\ w_2 &= \arctan\frac{\eta+\pi}{\zeta}, \quad -w_1 = \arctan\frac{\eta}{\zeta} \quad \text{if } \eta \leqslant 0. \end{aligned}$$

The expansion (9.3.2) is essentialy of Poincaré form, except for the prefactor which corresponds to the function $F^{\infty}$ in (9.2.81). Comparing (9.2.81) and (9.3.2) allows one to read off $F^{\infty}$ directly,

$$\begin{aligned} F^{\infty} = & \frac{\Gamma(1-b+\lambda)\Gamma(c)}{\Gamma(\frac{1}{2})\Gamma(c-b+\lambda)} 2^{a+b-1}(1-\mathrm{e}^{-\xi})^{1/2-c}(1+\mathrm{e}^{-\xi})^{c-a-b-1/2}\lambda^{-1/2}\times \\ & \left(\mathrm{e}^{(\lambda-b)\xi}+\mathrm{e}^{\mp i\pi(1/2-c)}\mathrm{e}^{-(\lambda+a)\xi}\right), \end{aligned} \tag{9.3.5}$$

with $a = b$ set to the appropriate value and $\lambda = \rho/2$. The higher terms in the Watson series (9.3.2) can in principle be obtained by retained higher order terms in the steepest descent calculation, but explicit formulae for them do not appear to be freely availably in the literature.

## 9.4. Spherically symmetric braneworlds

In this section we try and generalize the proof of Birkhoff's theorem to the braneworld. This is interesting because of the possibility the braneworld may support more general types of black hole than the Kerr or Kerr–Newman holes (Chandrasekhar, 1983) which are guaranteed to be the only candidates in four dimensions as a consequence of the Carter–Robinson theorem (Hawking and Ellis, 1973; Heusler, 1996). In particular, there is no five-dimensional uniqueness theorem (Emparan and Reall, 2002; Gutowski, 2004; Gutowski and Reall, 2004; Kodama, 2004; Reall, 2003), so black hole solutions which descend to the braneworld could conceiveably be quite strange (Kol and Wiseman, 2003; Kudoh and Wiseman, 2004). Kerr black holes are characterized by their conserved mass and angular momentum, and Kerr–Newman holes by these quantities plus the electric charge. However, one does not expect any new conserved quantities when moving to five dimensions, so the extra degrees of freedom should reasonably be expected to break the Carter–Robinson theorem. In view of this uncertainty, and the confusing observational situation it creates, the first step is to look at the consequences of the simplest assumption, that of spherical symmetry.

Black holes are interesting observational probes of strong-field gravity for a number of reasons. In the first case, the evidence for supermassive black holes which populate the central regions of almost all galaxies is now quite overwhelming, so conservative fears that black holes may not really exist in nature can be considerably allayed. Moreover, the extreme interest such supermassive objects generates means that observation efforts of quite remarkable skill and ingenuity are presently directed at recording their properties. Present experiments are becoming capable of probing regions fairly near the event horizon, so there is some hope of proper astronomical data with which to compare the predictions of any particular model. If the braneworld predicts a rather more liberal zoology of black holes than the Carter–Robinson theorem, then this might show up in the experimental data. As a second principal strand, the appearance of black holes in the early universe is

somewhat constrained (Guedens et al., 2002), and modifications of gravitational physics induced by brane world departures from conventional gravity at high energies might be probed using these bounds.

In four dimensions, it is known that the assumptions of spherical symmetry are already very strong, since the exterior asymptotically flat region of any spherically symmetric vacuum solution of Einstein's equations contains a timelike Killing vector.[3] This result is commonly known as Birkhoff's theorem, although the result itself had been conjectured earlier by Jebsen, in 1921. The commonly presented proof (Weinberg, 1994), though elementary, is somewhat turgid. An alternative approach (Schmidt, 1997) is to exploit the fact that when written in polar coordinates, a spherically symmetric $n$-dimensional flat space is itself a warped product, in the sense of Section 5.7.1, of a 2-sphere with an $(n-2)$-dimensional manifold. (This observation was spelt out explicitly in the footnote on p. 198.) In this section, we show, using the warped compactification method, that the Birkhoff theorem must suffer some modifications when passing to the braneworld. In the process we show how to reduce a general, spherically symmetric braneworld to three-dimensional Einstein gravity coupled to matter. This is based on the reduction in Schmidt (1997), but in the present derivation we adopt a more aesthetic approach of working entirely in the action, rather than using conformal transformation properties of the various curvature quantities.

We begin with a general warped product of $\mathbf{S}^2$ and a three-dimensional metric $\gamma_{ij}$ on a manifold $\mathcal{M}$,

$$\mathrm{d}s^2 = \gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j + \sigma^2(x)\mathrm{d}\Omega_2^2, \tag{9.4.1}$$

where $\mathrm{d}\Omega_2^2$ is the line element on a 2-sphere with coordinates $\theta$ and $\phi$, and indices $i$, $j$, ... label the three-dimensional coordinates. The connexion components are

$$\Gamma^{\theta}_{\phi\phi} = -\sin\theta\cos\theta \quad \Gamma^{\theta}_{\theta i} = \frac{\sigma_{,i}}{\sigma} \tag{9.4.2a}$$

$$\Gamma^{\phi}_{\phi\theta} = \cot\theta \quad \Gamma^{\phi}_{\phi i} = \frac{\sigma_{,i}}{\sigma} \tag{9.4.2b}$$

$$\Gamma^{n}_{ij} = \omega^{n}_{ij} \quad \Gamma^{n}_{\theta\theta} = -\gamma^{mn}\sigma\sigma_{,m} \quad \Gamma^{n}_{\phi\phi} = -\gamma^{mn}\sigma\sigma_{,m}\sin^2\theta, \tag{9.4.2c}$$

[3] One sometimes sees the formulation, "every spherically symmetric vacuum solution is static", but this is misleading because the Schwarzschild solution is not static inside the horizon.

where $\omega^n_{ij}$ is the three-dimensional metric connexion compatible with $\gamma_{ij}$. The curvature tensor satisfies

$$R_{ij} = \Omega_{ij} - \frac{2}{\sigma}\sigma_{,ij} + 2\omega^n_{ij}\frac{\sigma_{,n}}{\sigma} \tag{9.4.3a}$$

$$R_{\theta\theta} = 1 - -\gamma^{mn}{}_{,n}\sigma\sigma_{,m} - \gamma^{mn}\sigma\sigma_{,mn} - \gamma^{mn}\sigma_{,m}\sigma_{,n} - \gamma^{mn}\sigma\sigma_{,m}\omega^f_{nf}, \tag{9.4.3b}$$

and $R_{\phi\phi} = \sin^2\theta R_{\theta\theta}$ by symmetry. The quantity $\Omega_{ij}$ is the curvature built out out $\omega$. Taking the trace gives the Ricci scalar,

$$R = \gamma^{ij}\Omega_{ij} - \frac{4}{\sigma}\gamma^{ij}\sigma_{,ij} - \frac{2}{\sigma}\gamma^{ij}{}_{,j}\sigma_{,i} - \frac{2}{\sigma^2}\gamma^{ij}\sigma_{,i}\sigma_{,j} + \frac{2}{\sigma^2} + 2\gamma^j\omega^n_{ij}\frac{\sigma_{,n}}{\sigma} - 2\gamma^{ij}\omega^n_{jn}\frac{\sigma_{,i}}{\sigma} \tag{9.4.4}$$

One can use $R$ to construct the Einstein–Hilbert action, but this is most convenient when written in covariant form. This expression for $R$ can be covariantized using the rules

$$\nabla_i\nabla_j\sigma = \sigma_{,ij} - \omega^n_{ij}\sigma_{,n} \quad \text{so} \quad \sigma_{,ij} = \nabla_i\nabla_j + \omega^n_{ij}\sigma_{,n} \tag{9.4.5}$$

and

$$0 = \nabla_j\gamma^{ij} = \gamma^{ij}{}_{,j} + \omega^i_{jk}\gamma^{kj} + \omega^j_{jk}\gamma^{ik} \tag{9.4.6}$$

to replace $\sigma_{,ij}$ and $\gamma^{ij}{}_{,j}$ in the action. Taking account of possible surface contributions if the 3-dimensional manifold $\mathcal{M}$ has boundaries, that gives

$$S_{\mathrm{E}} = \frac{2\pi}{\kappa^2}\int_{\mathcal{M}} \mathrm{d}^3x\,\sqrt{-\gamma}\,\left(\sigma^2\gamma^{ij}\Omega_{ij} + 2\gamma^{ij}\nabla_i\sigma\nabla_j\sigma + 2\right) - \frac{8\pi}{\kappa^2}\int_{\partial\mathcal{M}} \mathrm{d}^2x\,\sqrt{-\theta}\,n^j\sigma\nabla_j\sigma, \tag{9.4.7}$$

in which $n_j$ is normal to $\partial\mathcal{M}$, and $\theta_{ij}$ is the induced metric on the boundary. Up to normalization, this is the action for three-dimensional dilaton gravity (Fujii and Maeda, 2003). In the braneworld, one should also include a Gibbons–Hawking term which accounts for the embedding of the brane in $\mathcal{M}$. It is easy to show that the trace of the extrinsic curvature must satisfy

$$\begin{aligned} K &= \theta^{ij}(\partial_i n_j - \Gamma^k_{ij}n_k) - \frac{1}{\sigma^2}\Gamma^i_{\theta\theta}n_i - \frac{\Gamma^i_{\phi\phi}}{\sigma^2\sin^2\theta}n_i \\ &= \operatorname{Tr}\tilde{K} + 2\frac{\gamma^{im}}{\sigma}\sigma_{,m}n_i, \end{aligned} \tag{9.4.8}$$

in which $\tilde{K}_{ij} = \nabla_i n_j$ is the second fundamental form in the three-dimensional submanifold. Thus, including all appropriate contributions, the total braneworld action should take the form

$$\begin{aligned} S =& \frac{2\pi}{\kappa^2}\int_{\mathcal{M}} \mathrm{d}^3x\,\sqrt{-\gamma}\,\left(\sigma^2\gamma^{ij}\Omega_{ij} + 2\gamma^{ij}\nabla_i\sigma\nabla_j\sigma + 2 - 2\Lambda\sigma^2\right) \\ &+ \frac{4\pi}{\kappa^2}\int_{\partial\mathcal{M}} \mathrm{d}^2x\,\sqrt{-\theta}\,\left(\sigma^2\operatorname{Tr}\tilde{K} + \lambda\sigma^2 + \kappa^2\mathcal{L}\sigma^2\right), \end{aligned} \tag{9.4.9}$$

where $\mathcal{L}$ is the Lagrangian for whatever matter and gauge theories are affixed to the brane.

After a fairly lengthy calculation, one recovers the field equations for $\sigma$ and $\gamma_{ij}$. They are

$$\Box\sigma = \sigma\left(\Lambda - \frac{\Omega}{2}\right) \tag{9.4.10a}$$

$$\Omega_{ij} - \frac{1}{2}\gamma_{ij}\Omega = -\frac{2}{\sigma^2}\left(\nabla_i\sigma\nabla_j\sigma - \frac{1}{2}\gamma_{ij}\nabla\sigma\cdot\nabla\sigma\right) + \gamma_{ij}\left(\frac{1}{\sigma^2} - \Lambda\right), \tag{9.4.10b}$$

showing that $\sigma$ is an unconventionally normalized scalar field, and that the expectation value $\langle\sigma^{-2}\rangle$ contributes to the local cosmological constant. The boundary conditions, from requiring that the action is stationary on $\partial\mathcal{M}$, are

$$n^j\nabla_j\sigma = -2\sigma(\tilde{K} + \lambda + \kappa^2\mathcal{L}) \tag{9.4.11a}$$

$$\frac{\sigma^2}{2}(\tilde{K}_{ij} - \tilde{K}\gamma_{ij}) = -\kappa^2\sigma^2\frac{\partial\mathcal{L}}{\partial\theta_{ij}} + \frac{1}{2}\gamma_{ij}\sigma^2(\lambda + \kappa^2\mathcal{L}) - \sigma n_{(j}\nabla_{i)}\sigma + \sigma\gamma_{ij}n^k\nabla_k\sigma. \tag{9.4.11b}$$

The route to the Birkhoff theorem involves a careful analysis of this system of equations in four dimensions. However, there is an important special case for which the Birkhoff theorem is trivial. This arises if $\sigma$ is a harmonic function on $\mathcal{M}$, so that $\Box\sigma = 0$. Eq. (9.4.10a) then shows that $\Omega = 2\Lambda$, which requires that $\mathcal{M}$ is a surface of constant curvature.

Let us make the Ansatz that $\gamma_{ij}$ takes the form of a braneworld compactification,

$$\mathrm{d}\gamma^2 = \theta_{\alpha\beta}\mathrm{d}x^\alpha\mathrm{d}x^\beta + \mathrm{d}y^2, \tag{9.4.12}$$

where $\partial\mathcal{M}$ is the surface $y = 0$. The interesting components of the curvature are

$$\Omega_{\alpha\beta} = \tilde{\Omega}_{\alpha\beta} - \frac{1}{2}\theta''_{\alpha\beta} - \frac{1}{4}\theta^{\rho\sigma}\theta'_{\alpha\beta}\theta'_{\rho\sigma} + \frac{1}{2}\theta^{\rho\sigma}\theta'_{\alpha\sigma}\theta'_{\beta\rho} \tag{9.4.13a}$$

$$\Omega_{yy} = -\frac{1}{2}\theta'^{\alpha\beta}\theta'_{\alpha\beta} - \frac{1}{2}\theta^{\alpha\beta}\theta''_{\alpha\beta} - \frac{1}{4}\theta^{\rho\sigma}\theta^{\alpha\beta}\theta'_{\alpha\rho}\theta'_{\beta\sigma}, \tag{9.4.13b}$$

where $\tilde{\Omega}$ is the curvature of $\theta$. The curvature scalar is, therefore,

$$\Omega = \tilde{\Omega} - \theta^{\alpha\beta}\theta''_{\alpha\beta} - \frac{1}{2}\theta'^{\alpha\beta}\theta'_{\alpha\beta} + \frac{1}{4}\theta^{\alpha\beta}\theta^{\rho\sigma}\theta'_{\alpha\sigma}\theta'_{\beta\rho} - \frac{1}{4}\theta^{\alpha\beta}\theta^{\rho\sigma}\theta'_{\alpha\beta}\theta'_{\rho\sigma}. \tag{9.4.14}$$

Carrying out the reduction of the bulk and boundary action in terms of $\theta$ yields a field equation for $\sigma$,

$$\tilde{\Box}\sigma = \frac{1}{2}\sigma(\tilde{\Omega} - \Theta) - \Lambda\sigma - \sigma'' - \frac{1}{2}\theta^{\alpha\beta}\theta'_{\alpha\beta}\sigma', \tag{9.4.15}$$

where $\Theta$ is the invariant combination

$$\Theta = \theta^{\alpha\beta}\theta''_{\alpha\beta} + \frac{1}{2}\theta'_{\alpha\beta}\theta'^{\alpha\beta} - \frac{1}{4}\theta^{\alpha\beta}\theta^{\rho\sigma}\theta'_{\alpha[\sigma}\theta'_{\beta]\rho}. \tag{9.4.16}$$

The field $\sigma$ is subject to a boundary condition

$$\sigma' = -\sigma(\lambda + \kappa^2\mathcal{L} - \theta^{\alpha\beta}\theta'_{\alpha\beta}), \tag{9.4.17}$$

at $y = 0$. For example, for a warped compactification, where $\theta_{\alpha\beta} = \mathrm{e}^{-2U(y)}\tilde{\theta}_{\alpha\beta}$, and assuming that $\sigma$ is harmonic in $\tilde{\theta}$, one has

$$\sigma'' - 2U\sigma' = \frac{1}{2}\sigma(\tilde{\Omega} - \Theta - 2\Lambda). \tag{9.4.18}$$

In particular because $\Theta$ is $y$-dependent, there is no possibility of setting $\sigma' = 0$ in order to obtain a 2-surface of constant curvature on the brane. This shows that the emergence of the Birkhoff theorem is not such a simple matter in the brane world.

APPENDIX A

# Functional analysis

Over the last several decades, modern methods in functional analysis have become increasingly important in theoretical and mathematical physics, and now supply important tools and methods in quantum mechanics and quantum field theory. Out of this collection of technologies, we will primarily be interested in two specific tools. A large part of the work described in later chapters makes use of the Sturm–Liouville transform, which is built out of approximations to a given function constructed from an infinite set of basis functions. We will also make heavy use of the path integral, which although still improperly understood from a rigorous standpoint has its basis in the theory of function spaces. In this appendix we supply the necessary background and give a concise introduction to the subject, without undue mathematical distraction. There is an extensive literature which can be consulted for more detailed explanations (Kolmogorov and Fomin, 1957; Riesz and Sz.-Nagy, 1955).

The classical function spaces are the Banach spaces $L^p$, which are normed spaces of measureable functions. Their study underpins most of the rigorous analysis we will undertake. Functions belonging to some $L^p$ for $p \geqslant 1$ have many important properties. We introduce Sturm–Liouville operators and the Sturm–Liouville eigenvalue equation, and study sets of eigenfunctions associated with these operators. This is sufficient to prove a classical theorem due to Rayleigh, that any $L^p$ function (for $p \geqslant 1$) on a compact set can be represented by a sum of Sturm–Liouville eigenfunctions. We generalize this result to the Sturm–Liouville transform. This transform has great general utility.

Although the notion of $L^p$-spaces underpins real analysis, there is no corresponding notion of a Lebesgue measure on $L^p$, which would allow a Lebesgue definition of integrals over $L^p$.[1] Nonetheless a heuristic definition can be given, essentially due to Dirac but

[1] Although this might seem like a mere mathematical technicality which is of no significance to physics or the natural world, this view is mistaken. Indeed, difficulties with a satisfactory definition of the functional measure give rise to anomalies in many quantum theories. These anomalies certainly impact on physics and our description of the world around us: anomalies are present in the Standard Model of particle

much refined by Feynman and many subsequent authors in the second half of the last century. This heuristic functional integral has many remarkable properties and is the basis for most modern work on quantum theory, where the approach provided by canonical quantization is difficult to apply, or needlessly burdensome. We shall use the path integral very extensively, both for detailed calculations and interpretations.

## A.1. The Liouville equation

The importance of the $L^p$ spaces lies in the control they provide over the behaviour of functions at infinity. As a result, functions belonging to some $L^p$ for $p \geqslant 1$ have a number of useful properties. Most of these properties can be related to the theory of self-adjoint operators, which are operators of the form $\mathcal{L}$ satisfying

$$\mathcal{L}u = \frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}u}{\mathrm{d}x}\right) + q(x)u = -\lambda\rho(x)u(x), \quad \text{where } x \in (a,b). \tag{A.1.1}$$

The functions $p(x)$ and $q(x)$ are arbitrary, but it is conventional to choose $p(x) > 0$ which can be accomplished if necessary by multiplying through by $-1$, and changing $\lambda \mapsto -\lambda$ to keep $\rho(x)$ positive. Since $p(x)$ is assumed continuous, it does not change sign over the interesting range $x \in (a,b)$. We make no assumptions about positivity of $q(x)$. The function $\rho$ is called the density, and $\lambda$ is known as the weight. We refer to $u(x)$ as an eigenfunction of weight $\lambda$.

The Sturm–Liouville problem is to find solutions $u(x)$ satisfying the boundary conditions. Typically these are taken to be mixed Dirichlet–Neumann conditions at $a$ and $b$, also sometimes called Robin boundary conditions,

$$\begin{aligned} \alpha u(a) + \alpha' \frac{\mathrm{d}u}{\mathrm{d}x}(a) &= 0 \\ \beta u(b) + \beta' \frac{\mathrm{d}u}{\mathrm{d}x}(b) &= 0 \end{aligned} \tag{A.1.2}$$

where the $\alpha$, $\alpha'$ and $\beta$, $\beta'$ are real constants, independent of $\lambda$. One finds that solutions do not exist for general choices for $\lambda$. Instead, Eq. (A.1.1) can be solved only for some specific values, which depend on the boundary conditions (A.1.2). One can study how the solutions $u_\lambda$ depend on the $\alpha$, $\alpha'$ and $\beta$, $\beta'$.

---

physics, although by luck (as it would seem) they are cancelled by the particular matter content this theory involves. For the purposes of this appendix, however, we are interested in functional integration as a purely mathematical exercise.

If $p(x) > 0$ everywhere, including the end-points, then the problem is said to be regular; if $p(x)$ vanishes at one or both end-points, then the problem is said to be singular. More generally, one says that the problem is singular if either $p$ or $q$ vanish at an end-point, or if the interval $(a, b)$ is unbounded. Although the regular Sturm–Liouville is simpler to study, most of our applications require the extension to the singular case.

**A.1.1. The regular Sturm–Liouville problem.** Let $L^2(a, b)$ be the Hilbert space of (possibly complex-valued) Lebesgue square-integrable functions on $(a, b)$. The natural setting for the Sturm–Liouville problem is the subspace $SL^2(a, b) \subset L^2(a, b)$ of square-integrable functions obeying the boundary conditions (A.1.2), which inherits a Hilbert space structure from $L^2(a, b)$. In this structure, the natural inner product is $(f, g) = \int_a^b \mathrm{d}x\, \overline{f} \cdot g$, where a bar denotes complex conjugation. However, the Liouville structure allows us to equip $SL^2(a, b)$ with another inner product. We denote this product $\langle f, g\rangle$ and define it as

$$\langle f, g\rangle = \int_a^b \mathrm{d}\mu(x) \bar{f} \cdot g \tag{A.1.3}$$

where $\mathrm{d}\mu(x) = \rho(x)\,\mathrm{d}x$ is called the Sturm–Liouville measure. The fundamental property of the Sturm–Liouville operator $\mathcal{L}$ is that it is self-adjoint in the $L^2$ inner product, meaning that $(f, \mathcal{L}g) = (\mathcal{L}f, g)$, as can be proved by integrating by parts twice and discarding a surface term which is zero because of the boundary conditions (A.1.2). The weights $\lambda$ are always real. This is an immediate consequence of the self-adjointness of $\mathcal{L}$. A similar argument shows that for distinct weights $\lambda_i$, $\lambda_j$, the eigenfunctions $u_i$, $u_j$ are orthogonal in the Sturm–Liouville inner product. To see this, write

$$(u_i, \mathcal{L}u_j) = -\lambda_j \langle u_i, u_j\rangle. \tag{A.1.4}$$

But by self-adjointness of $\mathcal{L}$, the left hand side is also equal to $-\lambda_i\langle u_i, u_j\rangle$ so if $\lambda_i \neq \lambda_j$ then $\langle u_i, u_j\rangle$ must be zero. In the case $i = j$ one can normalize the $\{u_i\}$ so that $\langle u_i, u_j\rangle = \delta_{ij}$. This normalization is frequently convenient.

**A.1.1.1. The Poincaré phase plane. Prüfer system.** One now wishes to examine which values of $\lambda_i$ are allowed, and how the $u_i$ behave as functions of $x$ and $\lambda_i$. The first step involves rewriting the Liouville equation in Prüfer form. This is a general transformation for second-order differential equations of the form

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(P(x)\frac{\mathrm{d}u}{\mathrm{d}x}\right) + Q(x)u = 0 \quad \text{where } x \in (a, b). \tag{A.1.5}$$

We assume $P(x) > 0$ but make no assumptions about $Q(x)$. One changes variables from $u$, $\mathrm{d}u/\mathrm{d}x$ to a new pair $r$, $\theta$ satisfying

$$\begin{aligned} P(x)\frac{\mathrm{d}u}{\mathrm{d}x} &= r(x)\cos\theta(x) \\ u(x) &= r(x)\sin\theta(x). \end{aligned} \tag{A.1.6}$$

This separates Eq. (A.1.5) into a pair of coupled first-order differential equations: an amplitude equation for $r$ and a phase equation for $\theta$

$$\begin{aligned} \frac{\mathrm{d}\theta}{\mathrm{d}x} &= Q\sin^2\theta + \frac{1}{P}\cos^2\theta \\ \frac{\mathrm{d}r}{\mathrm{d}x} &= \frac{1}{2}\left(\frac{1}{P} - Q\right) r\sin 2\theta. \end{aligned} \tag{A.1.7}$$

The great advantage of Prüfer's substitution is that it determines the oscillatory behaviour of $u$ in terms of a phase $\theta(x)$ which satisfies a first-order equation, independently of $r$. One can now apply rather straightforward existence and uniqueness theorems to the phase equation to reveal the qualitative behaviour of $u$ without having to solve the Liouville equation explicitly.

Let $\Theta$ be the initial phase at one boundary, say $x = a$. Therefore $\theta(a) = \Theta$. A solution to the phase equation exists for any $\Theta$, provided $P$ and $Q$ are continuous at $x = a$. The amplitude can then be found by quadrature,

$$r(x) = K\exp\int_a^x \frac{1}{2}\left(\frac{1}{P} - Q\right)\sin 2\theta\,\mathrm{d}x, \tag{A.1.8}$$

where $K$ is the initial amplitude, such that $r(a) = K$. Each solution to the Prüfer system depends on two constants, the initial phase $\Theta$ and the initial amplitude $K$. The zeroes of $u$ depend only on the phase, and occur for $\theta = 0, \pm\pi, \pm 2\pi, \ldots$. At these zeroes, $\theta$ is an increasing function of $x$, which follows immediately from the phase equation and the fact that $\sin^2\theta = 0$, $\cos^2\theta = 1$. Therefore, the zeroes of $u(x)$ are isolated, that is, are separated by a finite spacing in $x$. To see this, suppose that $x_n$ and $x_{n+1}$ are successive zeroes of $u(x)$. Therefore, $\theta(x_n) = n\pi$ and $\theta(x_{n+1}) = (n+1)\pi$. The derivative $\mathrm{d}\theta/\mathrm{d}x$ satisfies, at each point,

$$\left.\frac{\mathrm{d}\theta}{\mathrm{d}x}\right|_{x_n} = \frac{1}{P(x_n)}, \quad \text{and} \quad \left.\frac{\mathrm{d}\theta}{\mathrm{d}x}\right|_{x_{n+1}} = \frac{1}{P(x_{n+1})}, \tag{A.1.9}$$

which are bounded. Therefore one cannot let $x_n$ and $x_{n+1}$ become infinitesimally close without $\theta(x)$ becoming multivalued.

In fact, more is true. If $Q(x) > 0$ then $u(x)$ has exactly one extremum between two successive zeroes, for if $x_{1/2}$ is an extremum of $u(x)$, then $\mathrm{d}\theta/\mathrm{d}x|_{x_{1/2}} = Q(x_{1/2}) > 0$. Therefore $\theta(x)$ can cross the line $\theta(x) = (n+1/2)\pi$ only once, for if it did so twice then the slope would be negative the second time. One can say nothing about how many extrema occur if $Q < 0$.

**A.1.1.2. Liouville equation in Prüfer form.** The phase equation for the Sturm–Liouville problem is

$$\frac{\mathrm{d}\theta}{\mathrm{d}x} = (\lambda\rho + q)\sin^2\theta + \frac{1}{p}\cos^2\theta, \tag{A.1.10}$$

and the boundary conditions for the phase are

$$\tan\theta|_{x=a} = -\frac{\alpha'}{\alpha}\frac{1}{p(a)} = \tan\Delta \quad \text{and} \quad \tan\theta|_{x=b} = -\frac{\beta'}{\beta}\frac{1}{p(b)} = \tan\Gamma. \tag{A.1.11}$$

For convenience, we choose $\Delta$ and $\Gamma$ to satisfy $0 \leqslant \Delta < \pi$ and $0 < \Gamma \leqslant \pi$. Let $u$ be any solution of the Liouville equation, not necessarily satisfying the boundary conditions. If in addition one demands that $u$ obey the Dirichlet–Neumann conditions (A.1.2), then the phase $\theta$ must obey (A.1.11), except that one may have $\tan\theta(b) = \Gamma + n\pi$ for any integer $n \in \mathbf{Z}$. The boundary conditions (A.1.2) do not depend on $r$, so these conditions are both necessary and sufficient to Eq. (A.1.2) to hold.

We now have the following theorem,

**Theorem 1** (Oscillation theorem). The solution $\theta(x, \lambda)$ of the Sturm–Liouville–Prüfer differential equation (A.1.10) satisfying the initial condition $\theta(a, \lambda) = \Delta$ is a continuous, monotone increasing function of $\lambda$. Moreover, $\lim_{\lambda\to\infty}\theta(x, \lambda) = \infty$ (that is, $\theta(x, \lambda)$ is unbounded), and $\lim_{\lambda\to-\infty}\theta(x, \lambda) = 0$ at fixed $x$.

This fundamental theorem gives most of the important results in the subject as immediate corollaries.

Let $\theta(x, \lambda)$ be a solution of Eq. (A.1.10). It is clear that there is a smallest eigenvalue $\lambda$ such that $\theta(b, \lambda) = \Gamma$, that is, such that $\theta(x, \lambda)$ satisfies the Dirichlet–Neumann boundary conditions. This is true since $\theta(b, \lambda)$ increases without limit as $\lambda \to \infty$, starting from zero as $\lambda \to -\infty$. Similarly, there is an infinite sequence of eigenvalues $\lambda_n$ such that $\theta(b, \lambda_n) = \Gamma + n\pi$ for some integer $n \in \mathbf{Z}$. The sequence $\{\lambda_n\}$ is unbounded above. Thus, the Sturm–Liouville problem possesses an infinite set of eigenfunctions $\{u_n\}$, each of corresponding weight $\lambda_n$. Moreover, the $\{u_n\}$ each have exactly $n$ zeroes in the range $(a, b)$, not counting possible zeroes on the end-points themselves.

**A.1.1.3. Completeness of eigenfunctions. Rayleigh's theorem.** We now show that the $\{u_n\}$ for a complete basis for the Hilbert space $SL^2(a,b)$. Let $f \in SL^2(a,b)$ be any Lebesgue square-integrable function on $(a,b)$ satisfying the boundary conditions (A.1.2). We define the $n$th order residual approximant to $f$, written $f_n$, by

$$f_n = f - \sum_{i=0}^{n} \langle u_i, f \rangle u_i. \tag{A.1.12}$$

We assume that the $\{u_n\}$ are normalized so that $\langle u_n, u_m \rangle = \delta_{nm}$. Define a function space $W_n$ by $W_n = \text{span}\{u_i\}_{i=0}^n$. Then $f_n$ is orthogonal to $W_n$ in the Sturm–Liouville inner product. To see this, consider the inner product of $f_n$ with any element $w = \sum_i w_i u_i$ of $W_n$,

$$\begin{aligned} \langle f_n, \textstyle\sum_i w_i u_i \rangle &= \sum_i w_i \langle f, u_i \rangle - \sum_i \sum_j \overline{\langle u_j, f \rangle} w_i \langle u_j, u_i \rangle \\ &= 0. \end{aligned} \tag{A.1.13}$$

The spaces $W_n$ induce a family of polarizations of $SL^2(a,b)$. For each $W_n$, define $W_n^\perp$ by

$$SL^2(a,b) = W_n \oplus W_n^\perp, \tag{A.1.14}$$

which means that $W_n^\perp$ consists of those vectors in $SL^2(a,b)$ which are orthogonal to $W_n$. We now aim to prove that $W_n^\perp$ shrinks to zero, in an appropriate sense, as $n \to \infty$. This implies that any function in $SL^2$ can be written as a linear combination of elements in $W_\infty$.

Define the Rayleigh quotient for some test function $u \in SL^2(a,b)$ as

$$\mathcal{R}[u] = -\frac{(u, \mathcal{L}u)}{\langle u, u \rangle}. \tag{A.1.15}$$

This quantity was introduced by the English physicist and mathematician Lord Rayleigh[2]. Its utility lies in the useful property that $\mathcal{R}[u]$ satisfies a minimum principle when restricted to the $W_n^\perp$. In particular,

$$\lambda_{n+1} = \inf_{u \in W_n^\perp} \mathcal{R}[u], \tag{A.1.17}$$

[2]The Rayleigh quotient is more familiar in its finite-dimensional form, defined for a vector $\mathbf{x}$ and symmetric matrix $\mathbf{A}$ by

$$\mathcal{R}[\mathbf{x}] = -\frac{\mathbf{x}^T \cdot \mathbf{A} \cdot \mathbf{x}}{\|\mathbf{x}\|^2}. \tag{A.1.16}$$

To see this, write any non-zero vector $u \in W_n^{\perp}$ as $u = \sum_{i=n+1}^{\infty} a_i u_i$. Then $\mathcal{R}[u]$ satisfies

$$\mathcal{R}[u] = \frac{\sum_{i=n+1}^{\infty} \lambda_i |a_i|^2}{\sum_{i=n+1}^{\infty} |a_i|^2}. \tag{A.1.18}$$

Since $\lambda_{n+1}$ is the smallest remaining $\lambda_i$, it follows that $\inf_{u \in W_n^{\perp}} \mathcal{R}[u] = \lambda_{n+1}$. One can now show that as more and more terms are included in the approximation to $f$, the residual error $f_n$ decreases uniformly to zero.

**Theorem 2** (Rayleigh's theorem). As $n \to \infty$, $f_n \to 0$ almost everywhere. That is, $\|f_n\| = \langle f_n, f_n \rangle^{1/2} \to 0$.

Although this classical theorem is usually attributed to Rayleigh, an identical theorem in the context of Fourier analysis is known as Parseval's theorem. To prove Rayleigh's theorem, notice that since $f_n \perp W_n$, we have $\mathcal{R}[f_n] \geqslant \lambda_{n+1}$, since $f_n$, being a linear combination of elements of $SL^2(a,b)$, also lies in $SL^2(a,b)$. Then,

$$\begin{aligned} \lambda_{n+1} \|f_n\|^2 &\leqslant -(f_n, \mathcal{L} f_n) \\ &= -(f, \mathcal{L} f) - \sum_i \lambda_i |\langle u_i, f \rangle|^2, \end{aligned} \tag{A.1.19}$$

where we have used self-adjointness of $\mathcal{L}$ in the inner product, and the relation $(f, \mathcal{L} u_i) = -\lambda_i \langle f, u_i \rangle$. One now arranges, if necessary, for all the weights to be positive. This can always be achieved by reshuffling terms between the weight $\lambda$ and the auxiliary function $q$ in the Liouville equation. Thus,

$$\|f_n\|^2 \leqslant -\frac{(f, \mathcal{L} f)}{\lambda_{n+1}}. \tag{A.1.20}$$

The numerator is fixed, so taking the limit $n \to \infty$ gives $\|f_n\|^2 \to 0$, as was to be proved. One writes

$$f = \sum_{i=0}^{\infty} \langle u_i, f \rangle u_i \quad \text{almost everywhere,} \tag{A.1.21}$$

with the understanding, as in familiar Fourier analysis, that this representation may fail to be an equality at a finite number of points, where the right-hand side may exhibit finite discontinuities.

The proof of Rayleigh's theorem uses both important properties of the $\{\lambda_n\}$: that they form an infinite sequence, and that this sequence is unbounded above. In the absence of either of these conditions, one cannot obtain sufficient control over the approximant $f_n$ to

show that the error decreases uniformly to zero. Moreover, Eq. (A.1.21) is the starting point for our examination of the singular Sturm–Liouville problem.

**A.1.2. The singular Sturm–Liouville problem and the Sturm–Liouville transform.** Having brought the theory of the regular Sturm–Liouville problem to this point, we can take up again the case of the singular problem. The use of Sturm–Liouville theory later in this thesis will generally make use of the singular problem. For this reason, we should like to prove an analogue of the Rayleigh theorem, that functions $f \in SL^2(a,b)$ can be approximated arbitrarily closely be a series of eigenfunctions of the Liouville operator $\mathcal{L}$, even if one or both of $a$ or $b$ is infinite, or the functions $p$ has a zero at the end-points. The proof of the Rayleigh theorem does not directly use the regularity properties of the Sturm–Liouville operator, so our existing reasoning will carry over to the singular case provided we can establish the existence of an infinite sequence of weights $\{\lambda_n\}$ which is unbounded above. We will also need the orthogonality properties of the corresponding eigenfunctions $\{u_n\}$. Throughout, we assume that all functions are at least Lebesgue integrable on $(a,b)$, even if this interval is infinite.

The orthogonality property of the $\{u_n\}$ follows on integration by parts in the inner product $\langle f,g\rangle$, provided that the boundary term $[\bar{f}pg' - \bar{f}'pg]_b$ can be ignored. This is immediately true using the boundary conditions (A.1.2) in the regular case, but will also be true if $p = 0$ at the endpoint, provided the function $f$ and its derivative $f'$ are bounded there, or if $b = \infty$, provided both $p$ and $f$ fall off sufficiently fast at infinity. One may also employ periodic boundary conditions if the interval is finite, although we shall not make use of this possibility here.

To fix ideas, let us set $a$ to any finite value $a$, and allow $b$ to diverge to infinity. We return to the Liouville equation (A.1.1) and make the simultaneous change of variables (see Morse and Feshbach, 1953)

$$y = (p\rho)^{1/4}u, \quad \xi = \frac{1}{J}\int_a^x \sqrt{\frac{\rho}{p}}\,\mathrm{d}x, \quad \text{and} \quad J = \frac{1}{\pi}\int_a^b \sqrt{\frac{\rho}{p}}\,\mathrm{d}z. \tag{A.1.22}$$

This gives a transformed equation

$$\frac{\mathrm{d}^2y}{\mathrm{d}\xi^2} + \left[k^2 - w(\xi)\right] y = 0, \tag{A.1.23}$$

where $k^2 = J^2\lambda$ and $w(\xi)$ satisfies

$$w(\xi) = \frac{1}{(p\rho)^{1/4}} \frac{\mathrm{d}^2}{\mathrm{d}\xi^2} (p\rho)^{1/4} - J^2 \frac{q}{\rho}. \tag{A.1.24}$$

If $\lambda$ is large, then $k^2$ is large in comparison with $w(\xi)$, and the solution is approximately

$$u = \frac{1}{(p\rho)^{1/4}} \cos(k\xi + \theta), \tag{A.1.25}$$

where $\theta$ is some phase to be determined. Since $k^2$ is large, the derivative is dominated by the cosine rather than the prefactor $(p\rho)^{1/4}$, giving

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{(p\rho)^{1/4}} \frac{\mathrm{d}\xi}{\mathrm{d}x} \frac{\mathrm{d}}{\mathrm{d}\xi} \cos(k\xi + \theta) = -\frac{1}{(p\rho)^{1/4}} \frac{k}{J} \sqrt{\frac{\rho}{p}} \sin(k\xi + \theta). \tag{A.1.26}$$

Fitting the boundary conditions (A.1.2) gives the allowed, quantized values of $k$,

$$k_n \simeq n + \frac{J}{\pi n} \left( \frac{\beta}{\beta'} \sqrt{\frac{p}{\rho}}\bigg|_{x=b} - \frac{\alpha}{\alpha'} \sqrt{\frac{p}{\rho}}\bigg|_{x=a} \right). \tag{A.1.27}$$

Since $\lambda = (k/J)^2$, the weights $\lambda_n$ for large $n$ become approximately

$$\lambda_n \simeq \frac{n^2\pi^2}{\left( \int_a^b (\rho/p)^{1/2} \, \mathrm{d}x \right)^2}. \tag{A.1.28}$$

Therefore, as $b \to \infty$ one expects the spacing to go to zero and the discrete lattice of weights to approach a continuum, provided the combination $\rho/p$ behaves sensibly over the entire range. One can see this almost as easily through a weaker, informal argument: as $b \to \infty$ the boundary condition there is reduced to demanding that $f$ to go zero sufficiently fast at infinity. In this case any given $\lambda$ will satisfy the boundary conditions. For future convenience, we rescale $k$ so that $k_n = \sqrt{\lambda_n}$, to remove the normalizing factors of $J$. The asymptotic spacing between successive values of $k$ is $\pi / \int (\rho/p)^{1/2} \, \mathrm{d}x$, and the number of weights less than any given $k$ satisfies

$$n(k) \simeq \frac{k}{\pi} \int_a^b \sqrt{\frac{\rho}{p}} \, \mathrm{d}x. \tag{A.1.29}$$

Therefore the average number of weights falling between $k$ and $k + \mathrm{d}k$ becomes

$$\mathrm{d}n \simeq \left( \frac{1}{\pi} \int_a^b \sqrt{\frac{\rho}{p}} \, \mathrm{d}x \right) \mathrm{d}k. \tag{A.1.30}$$

The quantity in brackets amounts to the average density of weights for large $k$.

The most direct route to an analogue of the Rayleigh theorem now consists in considering the continuum limit of the normalization integral, $\langle u_m, u_n \rangle = \delta_{mn}$. As the diameter of $(a, b)$ diverges to infinity the weight lattice goes over to a continuum, and the discrete

labels $m$, $n$ become less useful. Instead we work with the weight label $k$, as in $\langle u_k, u_\ell \rangle$. As a function of the labels $k$, $\ell$, it is clear that $\langle u_k, u_\ell \rangle$ is zero whenever $k \neq \ell$, but we are still free to choose the normalization of the $\{u_k\}$. The most frequently useful normalization will be

$$\rho(k) \int_a^b \mathrm{d}x \; \rho(x) \overline{\psi_k}(x) \psi_\ell(x) = \delta(k - \ell). \tag{A.1.31}$$

This choice makes the transform/anti-transform pair symmetric. As a result, one has

$$f(x) = \int_{-\infty}^{\infty} \mathrm{d}k \; \rho(k) \psi_k(x) \int_a^b \mathrm{d}y \; \rho(y) \overline{\psi_k}(y) f(y). \tag{A.1.32}$$

We refer to this construction as the Sturm–Liouville transform. It is great general utility, and appears many times in the main chapters of this thesis.

One should not conclude that the singular Sturm–Liouville problem consists merely in trivial modifications to the regular case. This is not so. In general, the weight spectrum of a singular Sturm–Liouville problem may contain both discrete and continuous regions, and the weights may be unbounded above or below. The question of existence and convergence of Eq. (A.1.32) is not trivial. However, these technicalities will not concern us, since the Liouville operators $\mathcal{L}$ which will arise in our applications are sufficiently well-behaved that the theory as set out above is sufficient.

## A.2. The path integral

One might now aim to erect some theory of integration over $L^p(a, b)$ or at least $SL^p(a, b)$. Such a theory will be central to our presentation of quantum mechanics in later sections, but for the present we merely give an account of the mathematical elements of the theory. Unfortunately, there is no analogue of Lebesgue measure on $L^p(a, b)$, so one must proceed heuristically. Indeed, the construction of an appropriate functional measure is an outstanding problem in functional analysis, and a focus of continuous research. For the purposes of quantum mechanics, it is almost always sufficient to consider the restricted problem of integration on $SL^p(a, b)$. We will focus almost exclusively on this latter case.

**A.2.1. Integration on $SL^2(a, b)$.** Most of the functional integrals we shall need to consider are Gaussian, or can be reduced to sums of products of fields with a Gaussian factor. That is, the most general integral we shall need is of the form

$$I_n(x_1, \ldots, x_n) = \int_{SL^2(a,b)} [\mathrm{d}f] \; f(x_1) \cdots f(x_n) \exp\left( -\frac{1}{2} \int \mathrm{d}x \; f(x) \triangle f(x) \right), \tag{A.2.1}$$

where we have chosen to integrate over $SL^2(a,b)$, and $\triangle$ is a positive-definite regular Sturm–Liouville operator. Therefore $f$ can be expanded as a sum over the basis functions $\{u_n\}$ where each $u_n$ is a eigenfunction of $\triangle$ of weight $\lambda_n$[3]. This means that $f = \sum_n f_n u_n$ almost everywhere, where the expansion coefficients $f_n$ are real numbers. One can write the measure $[\mathrm{d}f]$ as a product of integrals over the $f_n$,

$$[\mathrm{d}f] = \Pi_n \mu \, \mathrm{d}f_n, \tag{A.2.2}$$

where we have kept open the possibility of an explicit normalization constant $\mu$. The remaining integrals are real, definite integrals over $\mathbf{R}$ which can be carried out explicitly. Thus, for example

$$I_0 = \prod_n \int \mu \, \mathrm{d}y_n \, \exp\left(-\frac{1}{2}\lambda_n f_n^2\right) = \prod_n \left(\frac{2\pi\mu^2}{\lambda_n}\right)^{1/2}. \tag{A.2.3}$$

For any matrix $\mathbf{A}$, the determinant $\det \mathbf{A}$ is defined as the product of the eigenvalues of $\mathbf{A}$, provided $\mathbf{A}$ is diagonalizable. By analogy we define the determinant $\det \triangle$ of the Liouville operator $\triangle$ as the product of its eigenvalues, that is, $\det \triangle = \prod_n \lambda_n$. Therefore,

$$I_0 = \left(\det \frac{\triangle}{2\pi\mu^2}\right)^{-1/2}. \tag{A.2.4}$$

Any determinant of this form diverges badly, since the weights $\lambda_n$ accumulate to infinity. To handle a divergence of this kind one must find some way truncate badly behaved quantities, such as Eq. (A.2.4), and render them finite. These so-called regularized expressions can then be worked with using normal mathematical procedures. One then attempts to rearrange the basic parameters of the theory in such a way that divergences do not reapppear on removing the regularization: this process is known as renormalization. We will shortly outline a modern renormalization scheme, known as $\zeta$-function regularization, which relies on analytical tools from complex variable theory. First, however, we describe how to extend the result (A.2.4) for $I_0$ to other $I_n$ with $n \geqslant 1$. The discussion given here largely follows that of Weinberg (1994).

[3]We have silently changed the sign of the weight $\lambda$, in comparison with our earlier definition of the Liouville operator $\mathcal{L}$. This sign change itself is of no real importance, but since the sign in the exponential Gaussian factor is of crucial importance in assessing convergence of the integral, we choose to make it explicit in (A.2.1) by making $\triangle$ positive definite.

Consider a generalized Gaussian integral of the form

$$\int [\mathrm{d}f] \exp\left(-\int \xi(x) f(x)\,\mathrm{d}x - \frac{1}{2}\int f(x)\triangle f(x)\,\mathrm{d}x\right) = \sum_{n=0}^{\infty} \frac{1}{(2n)!}\int \xi(x_1)\cdots\xi(x_{2n})\,\mathrm{d}^{2n}x\; I_{x_1\cdots x_{2n}}, \tag{A.2.5}$$

where we have expanded $\exp(-\int \xi f)$ in series and eliminated terms with an odd number of factors of $f$, which obviously give zero. On the other hand, one may evaluate the left-hand side by completing the square in the exponential and evaluating the resulting Gaussian integral exactly, using Eq. (A.2.4). Proceeding in this way, one finds

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{1}{(2n)!}\int \xi(x_1)\cdots\xi(x_{2n})\,\mathrm{d}^{2n}x\; I_{x_1\cdots x_{2n}} &= \left(\det\frac{\triangle}{2\pi\mu^2}\right)^{-1/2} \exp\left(\frac{1}{2}\int \mathrm{d}x\,\mathrm{d}y\; \xi(x)\triangle^{-1}(x,y)\xi(y)\right) \\ &= \left(\det\frac{\triangle}{2\pi\mu^2}\right)^{-1/2} \sum_{n=0}^{\infty}\frac{1}{2^n}\left(\int \mathrm{d}x\,\mathrm{d}y\; \xi(x)\triangle^{-1}(x,y)\xi(y)\right)^n\end{aligned} \tag{A.2.6}$$

Expanding the right hand side in powers of $\xi$ and comparing coefficients shows that $I_{x_1\cdots x_{2n}}$ must take the form of a sum of products of $\triangle^{-1}$. By symmetry, this must take the form

$$I_{x_1\cdots x_{2n}} \propto \sum_{\text{pairings}} \prod_{\text{pairs } (x,y)} \triangle^{-1}(x,y), \tag{A.2.7}$$

otherwise $I_{x_1\cdots x_{2n}}$ would not be invariant under exchange of indices. Taking careful account of numerical factors arising from the combinatorics, one finds

$$I_{x_1\cdots x_{2n}} = \left(\det\frac{\triangle}{2\pi\mu^2}\right)^{-1/2} \sum_{\text{pairings}} \prod_{\text{pairs } (x,y)} \triangle^{-1}(x,y). \tag{A.2.8}$$

This is commonly known as Wick's theorem.

**A.2.2. $\zeta$-function regularization.** It remains to explicitly evaluate the various functional determinants $\det\triangle$ which appear in Eq. (A.2.4) and Eq. (A.2.8). Such determinants are formally defined to be the product $\prod_n \lambda_n$ of the weights of $\triangle$. Throughout, we assume that $\triangle$ is self-adjoint in some appropriate inner product. In finite-dimensional terms, this is equivalent to demanding that $\triangle$ is Hermitian.

Instead of dealing directly with the divergent product $\prod_n \lambda_n$, one considers the $\zeta$-function

$$\zeta_{\triangle}(s) = \sum_n \lambda_n^{-s} \quad \text{where } s \in \mathbf{C}. \tag{A.2.9}$$

This sum is convergent provided $\mathrm{Re}(s)$ is sufficiently large. The choice of argument $s$ to denote a complex variable is conventional when dealing with $\zeta$-functions[4], even though $z$ is used universally elsewhere in complex analysis. We will drop the identifying $\triangle$ which signifies which operator $\zeta$ belongs to when no confusion can arise.

By differentiating $\zeta$, one obtains a simple formal expression for $\det \triangle$. Since $\zeta'$ must satisfy

$$\zeta'(s) = -\sum_n \lambda_n^{-s} \ln \lambda_n \tag{A.2.10}$$

it is clear that, at least formally, $\ln \det \triangle = -\zeta'(0)$. This is to be understood as defined by analytic continuation from the region of the $s$-plane where $\zeta(s)$ and $\zeta'(s)$ converge. Applying similar reasoning to the rescaled Liouville operator $\triangle$ gives an expression for $\det \triangle/2\pi\mu^2$ in terms of $\zeta(s)$,

$$\det \frac{\triangle}{2\pi\mu^2} = (2\pi\mu^2)^{\zeta(0)} \exp\left(-\zeta'(0)\right). \tag{A.2.11}$$

[4]The variable $s$ was used by Riemann in his initial work regarding the $\zeta$-function $\zeta_R(s) = \sum_m n^{-s}$, now known as the Riemann $\zeta$-function.

APPENDIX B

# Geometry and topology

Differential geometry, the mathematical study of surfaces and their properties, forms an indispensible part of modern theoretical physics. Differential geometry provides a common framework in which to express a large number of physical systems and endows the physicist with powerful tools with which to probe their behaviour. Once a given theory has been interpreted in terms of geometrical quantities, the analysis is often much facilitated by comparisons and analogies suggested by the formalism, which would otherwise have remained hidden. A large number of solved cases provided the mathematicians renders this strategy hugely expedient.

In this appendix, we briefly give an introduction to the basic tools of differential geometry, with the dual aim of fixing notation and providing a convenient reference for the main text. Regrettably a large number of incompatible sign conventions exist, which engenders a peculiar necessity to spell out, with some explicitness, which definitions and conventions are being used. We then pass to the theory of fibre bundles. The theory of bundles provides a sufficiently general framework to support all the technology described in the main text, and in particular, Einstein gravity, Yang–Mills theory, gauge fixing, and the description of warped compactifications. We recall the definition of parallel transport and define a connexion, which is used to construct a canonical derivative, known as the covariant derivative. This construction is most easily visualised in the context of bundle structures. We extend the theory to include anticommuting fields, which will become fermions in the physical theory, and describe the de Rham complex. Finally, we express some topological properties in terms of the de Rham cohomology and its complexification, the Dolbeault cohomology. These tools will be required for the discussion of string theory and brane physics to be given in Chapter 3, and for the work on Wess–Zumino models in the context of brane universes with compact, periodic extra dimensions described in Chapter 9. Throughout, we aim at brevity rather than an exhaustive treatment, and the reader is referred to the literature for further details (de Azcárraga and Izquierdo, 1995; de Felice

and Clarke, 1990; Deligne et al., 1999; Fuchs, 1992; Göckeler and Schücker, 1987; Green et al., 1987). In addition, there are numerous examples in the literature of the machinery of differential geometry applied to general relativity. These presentations are typically rather more economical in terms of the amount of mathematics which is developed. Particularly valuable instances are Chandrasekhar (1983); Hawking and Ellis (1973); Stewart (1991).

## B.1. Differential geometry

In this section, we repeat some elementary parts of tensor algebra and differential geometry, and apply them to gauge field theories and gravity. There are two approaches commonly found in the literature. One, often employed in elementary treatments, uses exclusively the Einstein holonomic gauge for tangent space indices[1]. In this model, the Riemann tensor $R^a{}_{bcd}$ and the covariant derivative $\nabla_a$ play a central role, and $R^a{}_{bcd}$ is interpreted as a measure of the curvature of the manifold. The other approach is based on the exterior algebra and avoids the early introduction of $\nabla_a$. The analogue of the curvature tensor $R^a{}_{bcd}$ arises via the Cartan structural equations. We prefer the Cartan approach for a number of reasons. It emphasizes the role of the connexion $A_a$ over $\nabla_a$, which makes the relationship with gauge field theory obvious. Moreover, it is the only way to introduce fermions into the theory. For example, the Cartan method is used in the construction of supergravities, where the Einstein gauge cannot be applied. This happens because in a holonomic gauge, the torsion must vanish. In this theories with fermions, the torsion is typically non-zero. In the exterior algebra formalism, the second Cartan structural equation conveniently summarises the role of torsion. Since we will use both gauge field theory and fermions, this method is preferable.

**B.1.1. Exterior algebra.** Let $V$ be a vector space of dimension $n$ over $\mathbf{R}$. The vector space $V^*$ dual to $V$ is the set of forms, that is, linear mappings $V \to \mathbf{R}$,

$$V^* = \{\phi \mid \phi : V \to \mathbf{R}\} \tag{B.1.1}$$

[1]By holonomic Einstein gauge, we mean the $GL(4)$ gauge in which the tangent space basis coincides with the holonomic coordinate basis $\partial/\partial x^a$ everywhere. This is not to be conufused with the weak-field gauge variously known as Einstein gauge, harmonic gauge, Fock gauge or de Donder gauge in which $\partial_a h^{ab} = 0$, where $h^{ab}$ is a small metric perturbation.

$V^*$ is a vector space of dimension $n$. Let $b_i$, $i = 1, \ldots, n$ be a basis of $V$. Then a basis $\beta^j$ of $V^*$, called the dual basis to $b_i$, is provided by defining $\beta^j(b_i) = \delta_i^j$, where $\delta_i^j$ is the Kronecker delta, equal to 1 when $i = j$ and zero otherwise.

Define $\Lambda^p$ to be the space of $p$-linear alternating forms,

$$\Lambda^p = \{\phi \mid \phi : \underbrace{V \times \cdots \times V}_{p \text{ times}} \to \mathbf{R}\} \tag{B.1.2}$$

By alternating, we mean that $\phi$ is antisymmetric under exchange of any two arguments. Evidently if $p > n$ then $\Lambda^p = \emptyset$. A form $\phi \in \Lambda^p$ is said to be of degree $p$. We define the wedge product $\phi \wedge \psi$ of a form $p$-form $\phi$ and a $q$-form $\psi$ to be the $(p+q)$-form

$$(\phi \wedge \psi)(v_1, \ldots v_{p+q}) = \frac{1}{p!q!} \sum_{\pi} \phi(v_{\pi(1)}, \ldots, v_{\pi(p)}) \psi(v_{\pi(p+1)}, \ldots, v_{\pi(p+q)}), \tag{B.1.3}$$

where $\pi$ is a permutation of $1, \ldots, p+q$. The wedge product is bilinear, associative and graded commutative,

$$\phi \wedge (a\psi + b\varphi) = a\phi \wedge \psi + b\phi \wedge \varphi \qquad \text{(bilinear)} \tag{B.1.4}$$

$$\phi \wedge (\psi \wedge \varphi) = (\phi \wedge \psi) \wedge \varphi \qquad \text{(associative)} \tag{B.1.5}$$

$$\phi \wedge \psi = (-1)^{pq} \psi \wedge \phi \qquad \text{(graded commutative)}, \tag{B.1.6}$$

where $a$ and $b$ are real; $\phi \in \Lambda^p$; and $\psi \in \Lambda^q$. The direct sum $\Lambda = \oplus_p \Lambda^p$, together with a product obeying these axioms is called a Grassman algebra or exterior algebra.

Owing to the antisymmetry of the wedge product, a basis of $\Lambda^p$ is provided by

$$\beta^{i_1} \wedge \cdots \wedge \beta^{i_p} \quad \text{where } 1 \leqslant i_1 < \cdots < i_p \leqslant n. \tag{B.1.7}$$

The restriction on the range of indices prevents overcounting. Therefore one can expand any $\phi \in \Lambda^p$ with respect to such a basis,

$$\phi = \phi_{i_1 \cdots i_p} \, \beta^{i_1} \wedge \cdots \wedge \beta^{i_p} \quad \text{where } 1 \leqslant i_1 < \cdots < i_p \leqslant n. \tag{B.1.8}$$

We adopt the usual Einstein summation convention, that any index appearing twice in super- and sub-script positions is to be summed over. Alternatively, one may allow the indices to range freely and divide by an appropriate symmetry factor,

$$\phi = \frac{1}{p!} \phi_{i_1 \cdots i_p} \, \beta^{i_1} \wedge \cdots \wedge \beta^{i_p}. \tag{B.1.9}$$

More generally, it can sometimes be convenient to allow forms $\phi$ to take values in some vector space $W$. This will occasionally happen in the sequel, but in this case the wedge

product cannot generally be defined unless some notion of multiplication exists in $W$, that is, $W$ is an algebra.

**B.1.2. Tangent vectors and differential forms.** Let $U$ be some open subset of $\mathbf{R}^n$. Consider any curve $q(t)$ on $U$. The tangent vector to $q(t)$ at $t = t_0$ is the operator $\partial/\partial t|_{t=t_0}$ which maps any function $f$ on $U$ to its derivative along $q(t)$, that is

$$\left.\frac{\partial}{\partial t}\right|_{t=t_0} f \mapsto \left.\frac{\partial f}{\partial t}\right|_{t=t_0} = \left( \left.\frac{\mathrm{d}x^a(q(t))}{\mathrm{d}t}\right|_{t=t_0} \frac{\partial}{\partial x^a} \right) f \tag{B.1.10}$$

where the $\{x^a\}$ are local coordinates on $U$. Therefore any tangent vector can be expressed as a linear combination of the coordinates $\partial/\partial x^a$. At a given $x \in U$ these basis vectors span a local copy of $\mathbf{R}^n$, known as the tangent space at $x$, and written $T_x(U)$. One may omit the argument $U$ where it is understood which coordinate patch $T_x$ is tangent to. We will often explot this convention to reduce clutter in formulae.

Let $\{b_\mu(x)\}_{\mu=1}^n$ be any set of $n$ linearly independent tangent vectors. We assume that the $b_\mu$ are smooth functions of position $x \in U$. Such a set is called a vielbein[2]. Any vector field $v(x)$ can be uniquely decomposed with respect to a given vielbein as $v(x) = v^\mu b_\mu(x)$. Notice that we label vielbein indices by Greek letters, whereas coordinate indices are labelled in Latin. If $\{c_\mu(x)\}_{i=1}^n$ is also a vielbein, then one must be expressible in terms of the other,

$$c_\mu(x) = \gamma_\mu{}^\nu b_\nu(x) \quad \text{where } \gamma \in GL(n) \tag{B.1.11}$$

and $GL(n)$ is the group of $(n \times n)$-dimensional real matrices. This property will be very important.

A differential form is any form $\phi(x)$ which acts on the tangent space at $x \in U$. One defines the wedge product of differential forms pointwise on $U$. Consider differential forms of degree one. Such differential forms live in the dual space to $T_x(U)$, called the cotangent space, and written $T_x^*(U)$. If $\{x^a\}$ are local coordinates on $U$, then $\{\partial/\partial x^a\}$ is a basis for $T_x$ and a dual basis is provided by the forms $\mathrm{d}x^a$,

$$\mathrm{d}x^a \left( \frac{\partial}{\partial x^b} \right) = \delta^a_b. \tag{B.1.12}$$

[2]Other names for this object are common in the literature. In four dimensions, one usually speaks of a tetrad or vierbein. We choose vielbein since the name is not dimension-dependent. Although the whole machinery of differential geometry can be satisfactorily developed without reference to a vielbein, it is a necessary construct for introducing spin-1/2 fields into the theory, as we shall wish to do.

(If desired, one can consider the derivative to act from the right, but this is not necessary.) The exterior derivative of a differential form is a map $\mathrm{d} : \Lambda^p \to \Lambda^{p+1}$; recall that the term derivative refers to the purely algebraic property of the Leibnitz rule. For functions $f(x)$ (that is, 0-forms) one defines

$$\mathrm{d}f = \frac{\partial f}{\partial x^a}\,\mathrm{d}x^a. \tag{B.1.13}$$

On general $p$-forms, the exterior derivative satisfies

$$\mathrm{d}\phi = \frac{1}{p!}\mathrm{d}\phi_{a_1\cdots a_p} \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_p}. \tag{B.1.14}$$

The exterior derivative satisfies a graded Leibnitz law,

$$\mathrm{d}(\phi \wedge \psi) = \mathrm{d}\phi \wedge \psi + (-1)^p \phi \wedge \mathrm{d}\psi \tag{B.1.15}$$

where $\phi$ and $\psi$ are, respectively, $p$- and $q$- forms on $U$. One can give a construction similar to a vielvein for a set $\{\beta^\mu\}$ of linearly independent 1-forms; such a set is called frame.

**B.1.3. Tensors.** Let us return to the tangent and cotangent spaces $T_x$ and $T_x^*$. Define a space $\Pi_q^p$ as the Cartesian product of tangent and cotangent spaces,

$$\Pi_q^p = \underbrace{T_x^* \times \cdots \times T_x^*}_{p \text{ times}} \times \underbrace{T_x \times \cdots \times T_x}_{q \text{ times}}. \tag{B.1.16}$$

A $(p, q)$-tensor, or tensor of rank $(p, q)$, is a function $T : \Pi_q^p \to \mathbf{R}$ from this space to the real numbers. (Or, more generally, one may allow arbitrary vector-space valued tensors, as with differential forms, but we do not make much use of this construction.) The space of all such functions is called a tensor product written

$$T_q^p = \underbrace{T_x \otimes \cdots \otimes T_x}_{p \text{ times}} \otimes \underbrace{T_x^* \otimes \cdots \otimes T_x}_{q \text{ times}}. \tag{B.1.17}$$

If $T \in T_q^p$ and $\beta^{i_1} \times \cdots \times \beta^{i_p} \times b_{j_1} \times \cdots \times b_{j_q}$ are vectors, then $T$ maps the $\beta^i$, $b_j$ to a real number $T(\beta^{i_1}, \ldots, \beta^{i_p}, b_{j_1}, \ldots, b_{j_q})$. One can define addition of tensors and multiplication of tensors by a scalar pointwise on $U$. With these conventions, $T_q^p$ is a vector space of dimension $n^{p+q}$ over $\mathbf{R}$, and the various tensor product spaces form an algebra over $\mathbf{R}$ under the multiplication $\otimes$. If $c_{i_1}, \ldots, c_{i_p}$ are vectors and $\omega^{j_1}, \ldots, \omega^{j_q}$ are one-forms then one also writes

$$c_{i_1} \otimes \cdots \otimes c_{i_p} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_q} \tag{B.1.18}$$

for the $(p, q)$-tensor which maps the $\beta^i$, $b_j$ to

$$\beta^{i_1}(c_{i_1}) \cdots \beta^{i_p}(c_{i_p}) \omega^{j_1}(b_{j_1}) \cdots \omega^{j_q}(b_{j_q}). \tag{B.1.19}$$

Usually the rank of a given tensor is clear from the context or made explicit by writing out components, so the qualifier $(p, q)$ is unnecessary and conventionally omitted.

**B.1.4. Push-forward and pull-back.** Let $U$ be an open subset of $\mathbf{R}^n$ with local coordinates $\{x^a\}$ and $V$ be an open subset of $\mathbf{R}^m$ with local coordinates $\{y^p\}$. Let $f$ be a mapping $f : U \to V$. Then $f$ induces a map of $T_x(U)$ to $T_{f(x)}(V)$, written $f_*$ and called the tangent mapping or $push - forward$ of $f$, as follows. Let $q(t)$ be a curve in $U$, so that $f(q(t))$ is a curve in $V$. The tangent mapping sends the tangent vector to $q(t)$ at $t = t_0$ to the tangent vector to $f(q(t))$ at $t = t_0$.

One can equally well define a pull-back of a differential form. If $\phi$ is a differential form on $V$, then one defines $f^*\phi$ to be a differential form on $U$ given by

$$f^*\phi(u) = \phi(f_* u), \quad \text{where } u \in T_x(U). \tag{B.1.20}$$

**B.1.5. Integration of $p$-forms.** The notion of integration makes use of the concept of an orientation. Consider some open subset $U$ of $\mathbf{R}^n$. An orientation on $U$ is a nowhere vanishing $n$-form $\omega$. Any frame $\{b_\mu\}$ on $\mu$ satisfying

$$\omega(b_1, \ldots, b_n) > 0 \quad \text{for all } x \in U. \tag{B.1.21}$$

is said to be oriented. Alternatively, one may work in terms of coordinate systems. Any set of local coordinates $\{x^a\}$ on $U$ is said to be oriented if the corresponding frame $\partial/\partial x^a$ is oriented,

$$\omega\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) > 0 \quad \text{for all } x \in U. \tag{B.1.22}$$

Conversely one may declare any set of local coordinates to be oriented by choosing the orientation form $\omega$ appropriately,

$$\omega = \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n. \tag{B.1.23}$$

Clearly an orientation is unique only up to multiplication by a positive definite function on $U$.

One defines integration for forms proportional to the orientation $\omega$. Let $\{x^a\}$ be an oriented coordinate system with orientation form $\omega = \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$. If $\phi = f(x)\omega$, then $\int_U \phi$ is defined to be

$$\int_U \phi = \int_U f(x)\, \mathrm{d}x^1 \cdots \mathrm{d}x^n \tag{B.1.24}$$

where the integral on the right-hand side is a conventional volume integral over $U$. The use of an orientation here is vital in defining the sign convention in the integral.

For $p$-form $\phi$, where $p < n$, let $K \subset U$ be a $p$-dimensional oriented surface in $K$ given by some embedding map $Q : T \mapsto U$ where $T$ is a coordinate on $\mathbf{R}^p$. Then one defines the integral of $\phi$ over $K$ as

$$\int_K \phi = \int_T Q^* \phi. \tag{B.1.25}$$

where $Q^*$ is the pull-back of $Q$, as above.

**B.1.6. Stokes theorem.** Most of the important classical integral theorems are special cases of the next theorem, conventionally known as Stokes' theorem, after the physicist and mathematician George Stokes. It includes, but is more general than, the theorem in elementary vector calculus which usually goes by the same name, that $\int_S \operatorname{curl} \mathbf{V} \cdot \mathrm{d}\mathbf{S} = \oint_C \mathbf{V} \cdot \mathrm{d}\boldsymbol{\ell}$, where the surface $S$ spans $C$, $\mathrm{d}\mathbf{S}$ is a vector element of area, and $\mathrm{d}\boldsymbol{\ell}$ is an element of length along $C$.

Stokes theorem says

$$\int_K \mathrm{d}\phi = \int_{\partial K} \phi \tag{B.1.26}$$

where $\phi$ is a $(p-1)$ form, and $K$ is a $p$-dimensional hypersurface.

## B.2. Metric structures

We now endow our open subsets of $\mathbf{R}^n$ with a metric. So far this step has been unnecessary; all the constructions given so far do not depend on the existence of a metric. However a metric is necessary for the formulation of general relativity, where it encodes details of the gravitational field in spacetime. (The formulation of general relativity in this way will be discussed later, after we have had a chance to introduce connexions and examine Yang–Mills theories.)

A metric on a vector space $V$ of dimension $n$ is a real, symmetric, non-degenerate bilinear form $g : V \times V \to \mathbf{R}$. Since the metric is bilinear, it is sufficient to know its values

on a basis $\{b_i\}_{i=1}^n$ of $V$. One defines the metric coefficients $g_{ij}$ by

$$g_{ij} = g(b_i, b_j). \tag{B.2.1}$$

Any metric admits a orthonormal basis, which is a basis $\{e_i\}_{i=1}^n$ such that $g(e_i, e_j) = \eta_{ij}$, where $\eta_{ij}$ is a diagonal matrix with an number $r$ of $+1$ entries and a number $s = n - r$ of $-1$ entries,

$$\eta_{ij} = \mathrm{diag}(\underbrace{+1, \ldots, +1}_{r \text{ times}}, \underbrace{-1, \ldots, -1}_{s = n - r \text{ times}}). \tag{B.2.2}$$

The number $s$ is called the signature of the metric. This is sometimes referred to as the Gram–Schmidt theorem. Notice that $\eta_{ij}$ cannot have any zero entries because $g$ is non-degenerate. One can also show that the integer $r$ does not depend on the choice of orthonormal basis; this is called Sylvester's Law of Inertia. If $r = n$, then the metric is called positive definite; if $r = 0$, the metric is negative definite. In any other case, the metric is called indefinite; this will be the case of physical interest.

Let $\{e_i\}_{i=1}^n$ be an orthonormal basis of $V$. Then $g$ defines a metric $g^*$ on $V^*$, called the dual or induced metric, as follows. Let $\{\varepsilon^i\}_{i=1}^n$ be the dual basis to $\{e_i\}$. Define $g^*$ to be orthonormal on $\{\varepsilon^i\}$, so that

$$g^*(e^i, e^j) = \eta_{ij}. \tag{B.2.3}$$

For a general basis $\{\beta^i\}$ of $V^*$, one defines

$$g^{ij} = g^*(\beta^i, \beta^j). \tag{B.2.4}$$

It is easy to see that $g^{ij}$ is the numerical inverse, in a matrix sense, of $g_{ij}$. Let $\{b_i\}$ and $\{\beta^j\}$ be dual, orthonormal bases of $V$ and $V^*$, which can be expressed in terms of orthonormal bases $\{e_\mu\}$ and $\{\varepsilon^\nu\}$,

$$b_i = b_i{}^\mu e_\mu \quad \text{and} \quad \beta^j = \beta^j{}_\nu \varepsilon^\nu, \tag{B.2.5}$$

where $b_i{}^\mu$ and $\beta^j{}_\nu$ are numerical matrices. The condition that $\{b_i\}$ and $\{\beta^j\}$ are dual requires that these matrices obey the reciprocity condition

$$\beta^i{}_\mu b_j{}^\mu = \delta^i_j. \tag{B.2.6}$$

Therefore,

$$g^{ij} g_{jk} = g^*(\beta^i, \beta^j) g(b_j, b_k) = g^*(\beta^i{}_\mu \varepsilon^\mu, \beta^j{}_\nu \varepsilon^\nu) g(b_j{}^\sigma e_\sigma, b_k{}^\tau e_\tau) = \delta^i_k \tag{B.2.7}$$

after using (B.2.6) and contracting indices. Therefore $\det(g_{ij})$ should be non-zero.

$V$ and $V^*$ are vector spaces of equal dimension, so they are isomorphic. Given a basis $\{b_i\}$ and its dual $\{\beta^j\}$ one can define

$$\begin{aligned} V &\cong V^* \\ b_i &\mapsto \beta^i. \end{aligned} \tag{B.2.8}$$

However, this isomorphism is not canonical. In the presence of a metric, one can define a canonical isomorphism by mapping an element $v \in V$ to the element $g(v, \cdot) \in V^*$. This is process is usually referred to as raising or lowering indices, since in components one has

$$v^i \mapsto g_{ji} v^i. \tag{B.2.9}$$

**B.2.1. Hodge star.** Consider the space of $p$-forms, $\Lambda^p$, over a vector space $V$. The dimension of $\Lambda^p$ and the dimension of $\Lambda^{n-p}$ match, so they are isomorphic:

$$\dim \Lambda^p = \dim \Lambda^{n-p} = \binom{n}{p}, \quad \text{so} \quad \Lambda^p \cong \Lambda^{n-p}. \tag{B.2.10}$$

Using a metric, one can define a canonical isomorphism. Let $\{e_i\}$ be a basis of $V$. Then if

$$\phi = \phi_{i_1 \cdots i_p} \, \mathrm{d}x^{i_1} \wedge \cdots \mathrm{d}x^{i_p} \tag{B.2.11}$$

is a $p$-form, one defines an $(n-p)$-form $*\phi$ by

$$*\phi = \frac{\sqrt{|\det g|}}{(n-p)!} \phi_{i_1 \cdots i_p} g^{i_1 j_1} \cdots g^{i_p j_p} \varepsilon_{j_1 \cdots j_n} \, \mathrm{d}x^{j_{p+1}} \wedge \cdots \wedge \mathrm{d}x^{j_n}, \tag{B.2.12}$$

where $\varepsilon_{j_1 \cdots j_n}$ is the totally antisymmetric rank $n$ tensor, sometimes called the Levi–Civita tensor, satisfying

$$\varepsilon_{12 \cdots n} = +1, \quad \varepsilon^{i_1 \cdots i_n} = g^{i_1 j_1} \cdots g^{i_n j_n} \varepsilon_{j_1 \cdots j_n} = \frac{1}{|\det g|} \varepsilon_{i_1 \cdots i_n}, \tag{B.2.13}$$

and $\det g$ is the determinant of the metric $g_{ij}$ on $V$. The operator $*$ is called the Hodge star, or sometimes the Poincaré star. The dual form $*\phi$ is called the Hodge or Poincaré dual, or (for brevity) just the dual form to $\phi$.

Let $U$ be a subset of $\mathbf{R}^n$. The Hodge dual of 1 is the invariant volume form on $U$,

$$*1 = \frac{\sqrt{|\det g|}}{n!} \varepsilon_{i_1 \cdots i_n} \, \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_n} = \sqrt{|\det g|} \, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n, \tag{B.2.14}$$

is called the invariant volume form on $U$.

The operation of dualising is an involution up to sign, that is, $*^2 = \pm 1$ with the sign depending details of the dimensionality $n$ of $V$, the signature $s$ of the metric (see Eq. (B.2.2)), and the rank $p$ of the form to which $*$ is applied,

$$*^2\phi = (-1)^{p(n-p)+s}\phi \quad \text{where } \phi \in \Lambda^p. \tag{B.2.15}$$

**Remark.** The Hodge star is the origin of a confusing construction in elementary vector algebra on $\mathbf{R}^3$ with a flat metric (that is, $g = \mathrm{diag}(1,1,1)$), where the cross product of two vectors $\mathbf{a}$, $\mathbf{b}$ is defined to be

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin\theta\, \hat{\mathbf{n}} \tag{B.2.16}$$

where $\theta$ is the angle between $\mathbf{a}$ and $\mathbf{b}$, and $\hat{\mathbf{n}}$ is a unit vector normal to the plane defined by $\mathbf{a}$ and $\mathbf{b}$. The cross product is sometimes written with as $\mathbf{a} \wedge \mathbf{b}$. This appears to give a natural algebraic structure on $\mathbf{R}^3$. However, this idea is false. Instead, this vector should more properly be viewed as the bivector found by exterior multiplication of $\mathbf{a}$ and $\mathbf{b}$,

$$a_i\, \mathrm{d}x^i \wedge b_j\, \mathrm{d}x^j = a_i b_j\, \mathrm{d}x^i \wedge \mathrm{d}x^j. \tag{B.2.17}$$

On dualising, this becomes

$$*(a_i\, \mathrm{d}x^i \wedge b_j\, \mathrm{d}x^j) = \sum_{i,j} \varepsilon_{ijk} a_i b_j\, \mathrm{d}x^k \tag{B.2.18}$$

which is the component expression for $\mathbf{a} \times \mathbf{b}$. There is no distinction between $V$ and $V^*$ in a flat metric; the components $a^i$ and $a_i$ are numerically equal. (We have written an explicit summation over $i$, $j$ in order to be explicit; summation over $k$ is implied, as usual.)

**B.2.1.1. Inner product of forms. Metric adjoint.** Using $*$ it is possible to construct an inner product on $\Lambda^p$. Let $\phi$ and $\psi$ be any two $p$-forms with support contained in $U$, so that $\phi \wedge *\psi$ is an $n$-form. The inner product $(\phi, \psi)$ of $\phi$ and $\psi$ is defined to be

$$(\phi, \psi) = \int_U \phi \wedge *\psi. \tag{B.2.19}$$

It can be verified, after a short calculation, this this is symmetric between $\phi$ and $\psi$.

The metric adjoint, $\delta$, of the exterior differential operator d is the adjoint of d in $(\cdot, \cdot)$. It satisfies

$$\delta = (-1)^{n(p+1)+s+1} * \mathrm{d} * . \tag{B.2.20}$$

Like d, the operator $\delta$ is nilpotent: $\delta^2 = 0$. $\delta$ is called the codifferential.

The sum of the differential and codifferential operators is called the Hodge–de Rham operator, $\not{d}$,

$$\not{d} = \mathrm{d} + \delta \tag{B.2.21}$$

The square $\not{d}^2$ is the Laplace–de Rham operator (or just Laplacian), $\triangle$. It is not too difficult to check that $\triangle$, acting on functions, is equivalent to (minus) the usual Laplacian,

$$\triangle f = \delta(\partial_k f \, \mathrm{d}x^k) = -\frac{1}{\sqrt{|\det g|}} \partial_k \left( \sqrt{|\det g|} g^{kl} \partial_l f \right). \tag{B.2.22}$$

The Laplace–de Rham operator is self-adjoint.

**B.2.2. Manifolds.** So far we have dealt entirely with coordinate regions $U$ which are are isomorphic to subsets of $\mathbf{R}^n$. A manifold of dimension $n$ is a topological space $M$ which locally looks like a subset of $\mathbf{R}^n$. To make this notion precise, consider an open subset $U$ of $M$, and a homeomorphism (that is, a one-to-one, invertible mapping) $\alpha$ of $U$ onto $\mathbf{R}^n$. Then $(U, \alpha)$ is called a chart of $M$ and $\alpha$ is a coordinate system. A manifold is a collection of charts $\{(U_i, \alpha_i)\}$ such that $\bigcup_i U_i = M$.

If the overlap region $U_i \cap U_j$ for any two charts is not zero, then the transition function $\alpha_j \circ \alpha_i^{-1}$ should be a sufficiently smooth function on $\mathbf{R}^n$. A manifold possessing such a structure is called a differentiable manifold.

Mostly it is possible to avoid explicitly dealing with manifolds, and work instead at the level of coordinate regions, as we have done so far. The existence of smooth transition functions $\alpha_j \circ \alpha_i^{-1}$ is usually sufficient to glue together constructions depending only on local conditions in neighbourhoods of $\mathbf{R}^n$ and apply them in the more general context of manifolds. However, in some cases, to address global questions it is occasionally necessary to refer to the more complicated structure as a whole, and not just from gluing together individual patches. This will be necessary in our later discussion of branes in string theory and cosmology. We use ideas from topology of manifolds when discussing Kaluza–Klein compactifications in Chapter 3 and Chapter 5, and when calculating the effects of string winding modes in Chapter 9.

## B.3. Homology and cohomology; the de Rham cohomology

Topological properties of manifolds, such as spacetime itself or the configuration space of gauge field theories, play an increasingly important role in physics. Homology is a branch of mathematics, initiated by Poincaré and later much refined by de Rham and Weil, which

measures topological properties of manifolds in an algebraic way, and provides an algebraic framework in which one can write equations and relations between toplogical features of a manifold.

We will focus particularly on the de Rham cohomology, because this is the concrete realization which is most useful in string theory. In particular, results in cohomology relate to the spectrum of massless or very light particles which appear after compactification to four dimensions. Topological properties of the compactification manifold determine the number and type of these particules, and such properties are preserved under purely metric deformations of the compactification manifold. In particular, the number of massless bosonic modes arising from a $p$-form in the higher dimensional theory is related to the $p$-th Betti number $b_p(M_6)$ of the manifold $M_6$. In our case, this will be useful in guaranteeing the existence of a zero mode in the four-dimensional gravity which arises from dimensional reduction of five-dimensional gravity in the bulk of a brane universe theory.

**B.3.1. de Rham cohomology.** Let $M$ be a manifold. A $p$ form $\phi$ on $M$ is called closed if $\mathrm{d}\phi = 0$. A $p$-form is called exact if there exists some $(p-1)$-form $\alpha$ such that $\phi = \mathrm{d}\alpha$ everywhere on $M$. The $(p-1)$-form $\alpha$ is then called a potential form for $\phi$.

It is easy to show that $\mathrm{d}^2 = 0$, that is, the result of applying d twice to any form is zero. Therefore, any exact form is closed. However, it may happen that a form be closed but not exact. The existence and number of closed forms which are not exact depends on the global, topological properties of $M$. For star-shaped regions of $\mathbf{R}^n$, a classical result due to Poincaré shows that there are no exact forms which are not closed.

**Theorem 3** (Poincaré's Lemma). Let $U$ be a contractible sub-manifold of $\mathbf{R}^n$. If $\mathrm{d}\phi = 0$ on $U$, then $\phi$ is exact.

Closed $p$-forms on $M$ are called de Rham $p$-cocycles; exact $p$-forms on $M$ are called de Rham $p$-coboundaries. For example, let $M$ be the circle $\mathbf{S}^1 = \{\phi \mid \phi \in (0, 2\pi]\}$. Consider the 1-form $\mathrm{d}\phi$. Despite appearances, this is not exact over $\mathbf{S}^1$, because $\phi$ is not a periodic function, and thus cannot be globally defined on $\mathbf{S}^1$. (Otherwise, by Stokes theorem, the perimeter of the circle would have length 0 rather than $2\pi$, since $\partial\mathbf{S}^1 = \emptyset$.)

The (real) de Rham cohomology groups $H^p_{DR}(M, \mathbf{R})$ over $M$ are the quotient spaces

$$H^p_{DR}(M, \mathbf{R}) = \frac{Z^p_{DR}(M, \mathbf{R})}{B^p_{DR}(M, \mathbf{R})}, \tag{B.3.1}$$

where $Z^p_{DR}(M, \mathbf{R})$ is the vector space of real-valued de Rham $p$-cocycles and $B^p_{DR}(M, \mathbf{R})$ is the vector space of real-valued de Rham $p$-coboundaries. The dimension of $H^p_{DR}(H, \mathbf{R})$ is called the $p$-th Betti number,

$$b_p(M) = \dim H^p_{DR}(M, \mathbf{R}). \tag{B.3.2}$$

The qualification real is made because of a subtler construction based on integer values. This construction is very important and plays a significant role in higher physics, but we shall not have much application for it in the work presented in this thesis. For this reason, we stick entirely to the real de Rham cohomology and usually drop both the qualification real and the argument $\mathbf{R}$ in $H^p_{DR}$, but occasionally the distinction is necessary.

The Betti numbers are topological invariants of $M$. Their alternating sum $\chi(M)$ is called the Euler–Poincaré invariant and is also a topological property,

$$\chi(M) = \sum_{p=0}^{n} (-1)^p b_p(M). \tag{B.3.3}$$

**B.3.2. General results for the de Rham cohomology.** There are no exact zero-forms, so $B^p_{DR}(M) = 0$. Also, a closed one-form is necessarily a constant, so $Z^0_{DR}(M) = \mathbf{R}$. Therefore, for any manifold, $H^p_{DR}(M) = \mathbf{R}$, provided $M$ is connected, and in general $H^0_{DR}(M) = \mathbf{R} \oplus \cdots \mathbf{R}$ with a separate copy of $\mathbf{R}$ for each connected component of $M$. For $\mathbf{R}^n$, Poincaré's Lemma establishes that $H^p_{DR}(\mathbf{R}^n) = 0$ for $p > 0$. Since this is a property of the fact that $\mathbf{R}^n$ can be covered by a single coordinate patch, a manifold $M$ will have non-trivial de Rham cohomology whenever the set of local coordinate neighbourhoods used to construct $M$ cannot be replaced by a single, global, coordinate chart.

The elements of $H^p_{DR}(M)$ define cohomology classes. If $\phi$ is a closed $p$-form then the equivalence class of $\phi$ in $H^p_{DR}(M)$ defines its cohomology class. Two closed $p$-forms $\phi$ and $\varphi$ define the same cohomology class if their difference is exact, that is, if $\phi - \varphi = \mathrm{d}\psi$ for some $(p-1)$-form $\psi$. For two such forms $\phi$ and $\psi$ defining the same cohomology class, their integrals over any closed submanifold $T$ of $M$ are equal,

$$\int_T \phi - \int_T \varphi = \int_T \mathrm{d}\psi = \int_{\partial T} \psi = 0, \tag{B.3.4}$$

since $\partial T = 0$. Thus, the integral of a closed form over a closed manifold depends only on its cohomology class. If $\phi$ is such that $\int_T \phi$ is an integer for all $T$, then $\phi$ is said to define an integral cohomology class.

**B.3.3. de Rham cohomology ring.** The wedge product on $M$, for the real cohomology, induces a ring structure on the $H^p_{DR}(M)$. Let $\phi$ and $\psi$ be, respectively, closed $p$- and $q$-forms on $M$. Then their wedge product $\phi \wedge \psi$ is also closed, in virtue of the graded Leibnitz property (B.1.15). Moreover, the cohomology class of $\phi \wedge \psi$ can be calculated from the cohomology classes of $\phi$ and $\psi$, for if one replaces $\phi$ by an equivalent (in cohomology) form $\phi + \mathrm{d}\alpha$, one has

$$(\phi + \mathrm{d}\alpha) \wedge \psi = \phi \wedge \psi + \mathrm{d}\alpha \wedge \psi = \phi \wedge \psi + \mathrm{d}(\alpha \wedge \psi) \tag{B.3.5}$$

in virtue of the closedness of $\psi$. Given representative elements $\phi$, $\psi$ of cohomology classes $H^p_{DR}(M)$, $H^q_{DR}(M)$, their wedge product uniquely defines a cohomology class in $H^{p+q}_{DR}(M)$. This defines what is called the cohomology ring of $M$.

**B.3.4. Simplicial homology.** Now consider any series $c_i$ of closed $p$-dimensional surfaces in $M$. A $q$-chain in a formal sum $\lambda^i c_i$ with real-valued coefficients $\lambda_i$. A $q$-cycle is a $q$-chain without boundary, $\partial c_q = 0$, and a $q$-boundary is a $q$-cycle which is the boundary of a $(q+1)$-chain, $c_q = \partial c_{q+1}$. The (real) simplicial homology of $M$ is given by

$$H_p(M) = \frac{Z_p(M)}{B_p(M)}, \tag{B.3.6}$$

where $Z_p(M)$ is the vector space of $q$-cycles and $B_p(M)$ is the vector space of $q$-boundaries. The product of a cycle $c_p \in Z_p(M)$ and a form $\phi \in Z^p_{DR}(M)$ is given by

$$\langle c_p \mid \phi \rangle = \int_{c_p} \phi. \tag{B.3.7}$$

The product $\langle c_p \mid \phi \rangle$ is called a period. Using this formalism, Stokes' theorem reads $\langle c \mid \mathrm{d}\phi \rangle = \langle \partial c \mid \phi \rangle$, in which case $\partial$ is the adjoint of d in $\langle \cdot \mid \cdot \rangle$, and vice-versa.

The simplicial homology and the de Rham cohomology and formally dual to each other, via the inner product $\langle \ \mid \ \rangle$. Notice that this is not a metric: one cannot use $\langle \ \mid \ \rangle$ to raise or lower indices on a homology $q$-cycle in order to transform into into a cohomology $q$-class, or vice versa. Nonetheless, the analogy with forms and vectors in the absence of a metric structure is profitable. Homology is a simpler algebraic structure than cohomology; it provides a context in which one can write relations between homology cycles, which is essentially determined by the laws of contour integration. Consider the surface shown in Fig. B.1. This surface contains three homology 1-cycles. Since it is two-dimensional, there

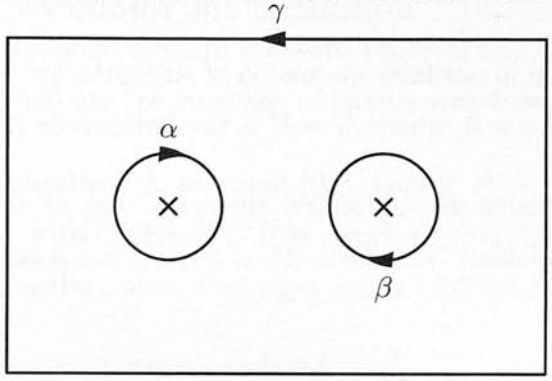



**Figure B.1.** Toplogically non-trivial surface with three homology cycles, here labelled $\alpha$, $\beta$ and $\gamma$. Each cross is a puncture on the manifold where a point has been removed.

are no higher cycles. We denote these cycles $\int_\alpha$, $\int_\beta$ and $\int_\gamma$. One can write (in homology)

$$\int_\gamma = \int_\alpha + \int_\beta . \tag{B.3.8}$$

This follows from the general discussion of simplicial homology given above, but it is easy to see that it follows on the basis of the usual rules of integration, provided integrals along interior arcs can be neglected. This motivates the notation for homology cycles as integrals.

However, this is the limit to which homology can be endowed with an algebraic structure. As we have discussed above, the dual structure — cohomology — has much richer algebraic properties. In particular, it can be endowed with a graded ring structure. At the purely technical level, this makes cohomology both easier to work with and more powerful than homology, in virtue of the deeper structure and larger array of tools available to probe it. The cohomology classes on the manifold of Fig. B.1 can be written almost everywhere as

$$\frac{1}{2\pi}\mathrm{d}\theta_\alpha, \quad \frac{1}{2\pi}\mathrm{d}\theta_\beta, \quad \text{and} \quad \frac{1}{2\pi}\mathrm{d}\theta_\gamma \tag{B.3.9}$$

where $\theta_\alpha$, $\theta_\beta$ are polar coordinate describing the cycles $\int_\alpha$, $\int_\beta$ (respectively), which are well-defined locally in a coordinate patch surrounding each puncture, and $\mathrm{d}\theta_\gamma$ cannot easily be written as a local exterior derivative, but integrates to $2\pi$ along $\gamma$ and to zero along $\alpha$ and $\beta$.

**B.3.5. Hodge–de Rham theory.** There is an important application of the de Rham cohomology groups to the Hodge–de Rham operator $\not{d}$ introduced in Section B.2.1.1 above,

which will prove vital in our later discussion of string theory. A differential form $\phi \in \Lambda^p$ is called harmonic if $\triangle\phi = 0$, where $\triangle = \not{d}^2$ is the Laplace–de Rham operator.

**Theorem 4** (Hodge decomposition theorem)**.** Let $M$ be a compact, oriented Riemannian manifold without boundary. Then any $p$-form $\phi$ on $M$ admits a unique, global decomposition,

$$\phi = \mathrm{d}\omega + \delta\beta + \gamma, \tag{B.3.10}$$

where $\gamma$ is harmonic. Moreover, $(\mathrm{d}\omega, \delta\beta) = (\mathrm{d}\omega, \gamma) = (\delta\beta, \gamma) = 0$.

This theorem is sometimes summarised by saying that any form can be written as the sum of an exact form, a coexact form, and a harmonic form. Notice that if $\phi$ is exact and coexact, then $\triangle\phi = 0$, so $\phi$ is harmonic. The converse is also true if $M$ is a compact, boundaryless Riemannian manifold, for in that case $(\triangle\phi, \phi) = (\delta\phi, \delta\phi) + (\mathrm{d}\phi, \mathrm{d}\phi)$, so if $\triangle\phi = 0$ then both $\mathrm{d}\phi$ and $\delta\phi$ must separately by zero. Therefore, let $\phi$ be closed. Then the term $\delta\beta$ involving the codifferential must be absent in the Hodge decomposition of $\phi$, so that $\phi = \mathrm{d}\omega + \gamma$. This means that the harmonic form $\gamma$ and $\phi$ define the same cohomology class; each de Rham cohomology class contains a harmonic representative. In fact, more is true.

**Theorem 5.** On a compact, orientable, boundaryless Riemannian manifold $M$, $H^p_{DR}(M)$ is isomorphic to the space $\mathrm{Harm}^p(M)$ of harmonic $p$-forms.

$$\dim \mathrm{Harm}^p(M) = \dim H^p_{DR}(M) = b_p, \tag{B.3.11}$$

where $b_p$ is the $p$-th Betti number of $M$.

## B.4. A summary of fibre bundles

Fibre bundles play an important unifying role in modern physics. On the one hand, they provide a sufficient mathematical superstructure in which to frame almost all physical theories of contemporary interest, including, in particular, general relativity, supergravity, and gauge theory. Fibre bundles also have applications to quantization and quantum theory, which we will take up in Section 2.1. One the other hand, it is the notion of fibre bundles that lends the geometrical formulation of physics much of its elegance and sophistication. A large number of mathematical tools and constructions which can be applied to physical theories in a geometrical sense find their most natural expression in

terms of fibre bundles. Therefore, there are both physical and mathematical reasons to introduce bundles. We shall use the language of bundle structures very extensively.

**B.4.1. Principal bundles.** A principal fibre bundle $P(G, M)$ over a manifold $M$, called the base manifold, with Lie group $G$ (the structure group), consists of a manifold $P$ (the total manifold) and a (by convention) right action of $G$ on $P$, $P \times G \to P$, such that

(1) the action of $G$,

$$g : p \mapsto pg \equiv R_g p \tag{B.4.1}$$

is free, which means that if $p = pg$ then $g$ is the unit element $e$ of $G$.

(2) $M$ is the quotient space $P/\mathcal{R}$, where $\mathcal{R}$ is the equivalence relation induced by the action of $G$. The differentiable mapping $\pi : P \to M$ which takes equivalence classes of points in $P$ to the base manifold $M$ is called the projection. Acting on points with $g$ does not change the projection; $\pi(xg) = \pi(x)$. The group $G$ is often called the structure group of the bundle.

(3) $P$ is locally trivial, in the sense that every point $x \in M$ has a local neighbourhood $U$ such that $\pi^{-1}(U)$ is diffeomorphic to $U \times G$ (with $G$ considered as a manifold).

The inverse image $\pi^{-1}(x)$ of a point $x \in M$ is called the fibre over $x$. Each fibre is stable under the action of $G$; fibres are isomorphic to individual copies of $G$.

A local cross section over an open set $U \subset M$ is a differentiable mapping $\sigma : U \to P$ such that $(\pi \circ \sigma)(x) = x$ for all $x \in U$. Therefore a local cross section is a mapping of each point in $U$ to a unique point in the bundle.

Many important concepts in physics and mathematics turn out to be examples of principal bundles, although other types of bundles do exist which we shall examine later. We give two important examples of principal bundles: Lie groups, and bundles over spheres.

**B.4.1.1. Lie groups.** Let $G$ be a Lie group, and $K$ a closed subgroup. $K$ acts freely on $G$ by right translations, ie., $k : g \in G \mapsto gk \in G$. The factor space $G/K$ contains the equivalence classes $\{g\}$ up to translation by $K$. Then $G(K, G/K)$ is the principal bundle over the quotient $G/K$.

**Remark.** A very important example in this category is the Hopf fibration. The Hopf bundle is the principal bundle $SU(2)(U(1), \mathbf{S}^2)$, where $\mathbf{S}^2 = SU(2)/U(1)$ is a two-sphere. To see how this arises, consider the Lie group $SU(2)$. As a classical matrix group, it is the Lie group of unitary $2 \times 2$ matrices with determinant unity. Therefore, the elements

of $SU(2)$ can be represented by matrices $g$ of the form

$$g = \begin{pmatrix} z_1 & -z_2^* \\ z_2 & z_1^* \end{pmatrix} \quad \text{where } z_1 z_1^* + z_2 z_2^* = 1 \text{ and } z_1,\, z_2 \in \mathbf{C}. \tag{B.4.2}$$

If one writes out the real and imaginary components of the $z_i$ explicitly, $z_i = x_i + \mathrm{i}y_i$, then the condition on the determinant of $g$ becomes

$$x_1^2 + y_1^2 + x_2^2 + y_2^2 = 1. \tag{B.4.3}$$

This is the equation for a 3-sphere $\mathbf{S}^3$, and is the total space of the Hopf bundle. The structure group, or fibre, is $U(1)$. In the same representation, this is

$$g_0 = \begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix} \tag{B.4.4}$$

and the right action of $U(1)$ on $SU(2)$ is simply

$$g \mapsto g g_0 = \begin{pmatrix} z_1\zeta & -z_2^*\zeta^* \\ z_2\zeta & z_1^*\zeta^* \end{pmatrix}. \tag{B.4.5}$$

The projection $\pi : \mathbf{S}^3 \to \mathbf{S}^2$ is the so-called Hopf map,

$$\pi(z_1, z_2) \mapsto \hat{\mathbf{x}} = \left[z_2^* z_1 + z_1^* z_2, \mathrm{i}(z_2^* z_1 - z_2 z_1^*), |z_1|^2 - |z_2|^2\right] \tag{B.4.6}$$

where each of the components on the right hand side is real, as can be verified by writing each expression explicitly out in components. Since $|z_1|^2 + |z_2|^2 = 1$, it follows that $|\hat{\mathbf{x}}| = 1$, so that the right-hand side does indeed parametrize $\mathbf{S}^2$.

This is not the only possible Hopf fibring of a sphere, although the possibilities are restricted, lying in fact in a one-to-one relationship with the possible division algebras. The possible Hopf fibrings are $\mathbf{S}^1 \to \mathbf{S}^3 \to \mathbf{S}^2 = CP^1$, as studied above, where $CP^1$ is the complex projective space; $\mathbf{S}^3 \to \mathbf{S}^7 \to \mathbf{S}^4 = HP^1$; and $\mathbf{S}^7 \to \mathbf{S}^15 \to \mathbf{S}^8 = OP^1$, where $HP^1$ and $OP^1$ are, respectively, the quaternionic and octonionic projective spaces. To make the full complement of division algebras one can add the (trivial) fibring $\mathbf{S}^0 \to \mathbf{S}^1 \to RP^1$. The spheres $\mathbf{S}^0$, $\mathbf{S}^1$, $\mathbf{S}^3$ and $\mathbf{S}^7$ can be identified with the unit real numbers, complex numbers, quaternions and octonions, respectively.

**B.4.2. Bundles of other types.** A vector bundle is a fibre bundle in the sense of the previous section, where the typical fibre is some vector space $F$. Many of the bundles we will encounter shall be of this sort; if the vector space $F$ is of dimension one, then the bundle is usually referred to specially as a line bundle.

The most important example of a vector bundle is the tangent bundle of some manifold $M$. In this case the total manifold $T(M) = \bigcup_x T_x(M)$ is the union of all the tangent spaces to $M$. The projection is chosen to send any $X(x) \in T_x(M)$ to the base point $x \in M$, and the fibre is the vector space $T_x(M) \cong \mathbf{R}^n$. One can similarly define the cotangent bundle and general tensor bundles.

Let $P(G, M)$ be a principal bundle, and $F$ a vector space on which $G$ has a left action,

$$g : v \in F \mapsto gv. \tag{B.4.7}$$

Define a right action of $G$ on $P \times F$ by

$$g : (p, v) \mapsto (p, v)g = (pg, g^{-1}v). \tag{B.4.8}$$

This action determines an equivalence relation $\mathcal{R}$ on $P \times F$, given by elements which are related by orbits of $G$,

$$(p', v')\mathcal{R}(p, v) \quad \text{iff there exists some } g \text{ so that} \quad (p, v)g = (p', v'). \tag{B.4.9}$$

The set of equivalence classes $\{p, v\}$ is denoted $E$. The associated bundle to $P(G, M)$ is defined by taking the total manifold to be $E$, the base space to be $M$, and the standard fibre to be $F$ with structure group $G$. The projection $\pi_E$ is defined to take an equivalence class $\{p, v\}$ to the point $\pi(p) \in M$, where $\pi$ is the projection in the principal bundle $P(G, M)$.

## B.5. Connexions

Let us now return to physics. For example, in field theory matter fields are described by sections of associated bundles, whereas Yang–Mills fields are described by connexions on principal bundles. General relativity is described by a bundle of linear frames over spacetime, which describes coordinate systems in which special relativity applies. The extra structure one needs to define any of these, in addition to the bundle itself, is a connexion.

Let $P(G, M)$ by a bundle over $M$, and consider any path $q(t)$ on an open subset $U$ of $M$ together with a local cross section $\sigma : U \to P$. As one traverses $q$ the cross section $\sigma(q(t))$ takes values in adjacent fibres. These fibres are all isomorphic to the standard fibre $P/G$, but not canonically, so there is no way to directly compare $\sigma(q(t))$ to the adjacent value $\sigma(q(t+\varepsilon))$. The bundle itself does not provide any mechanism to achieve this. Instead, one must add extra elements; the extra element needed to enable the comparison is called a connexion.

**B.5.1. Parallel transport.** To define a connexion, it is most convenient to begin at the level of paths $q(t)$ on open subsets of $M$; later, we will descend to the infinitesimal level where the mathematics of the description is easier to handle. For the present however, descriptions are easier in integrated form. Let $q(t)$ be a path between points $q_x$ and $q_y$, and let $\Gamma_{[}q]$ be an element of $G$ which depends on the path. $\Gamma[q]$ is called the parallel transporter along $q(t)$. To compare $\sigma(q_x)$ with $\sigma(q_y)$, considered as elements of the standard fibre $P/G$, we pick any isomorphism of the fibres at $q_x$, $q_y$ onto $P/G$. For example, in the case where the bundle is a vector bundle with standard fibre as the vector space $F$, then this is equivalent to picking a basis at $q_x$ and $q_y$, resulting in different copies $F_x$ and $F_y$ of $F$. The result of letting $\Gamma[q]$ act on $\sigma(q_x) \in F_x$ is considered to lie in $F_y$. $\Gamma[q]\sigma(q_x)$ and $\sigma(q_y)$ can then be directly compared, since they lie in the same vector space.

For this to make sense, the result must be gauge invariant: that is, it must be invariant under $G$-transformations of the fibres at $q_x$ and $q_y$, or, in the vector bundle analogy, it must be independent of the basis one picks for $F_x$, $F_y$. Let $\gamma(x)$ be a $G$-valued function on $U$. The result of acting with $\gamma$ should commute with $\Gamma$, so the result of making the transformation and then parallel transporting

$$\Gamma'[q]\gamma(q_x)\sigma(q_x) \tag{B.5.1}$$

should be the same as parallel transporting and then making the transformation,

$$\gamma(q_y)\Gamma[q]\sigma(q_x). \tag{B.5.2}$$

The symbol $\Gamma'$ denotes the gauge transformed parallel transporter. Equating these two as elements of $G$ shows that

$$\Gamma'[q] = \gamma(q_y)\Gamma[q]\gamma^{-1}(q_x). \tag{B.5.3}$$

This is the $G$-transformation law for parallel transport. Transformations like $\gamma(x)$ are usually known as gauge transformations in physics. In this language, (B.5.3) records how the parallel transporter responds to gauge transformations.

Now let $q(t)$ be an infinitesimal curve at $q_x$. Such a curve is defined by its tangent vector $X$, and since parallel transport is continuous by assumption, $\Gamma[q]$ must be very close to the identity in $G$,

$$\Gamma[q] = \mathbf{1} + \omega(X) \tag{B.5.4}$$

where $\omega(X)$ is the result of operating on $X$ with a $\mathfrak{g}$-valued 1-form $\omega_a$,[3] and $\mathbf{1}$ is the unit element of $G$ (or $\mathfrak{g}$). The $\mathfrak{g}$-valued form $\omega_a$ is called the connexion, and the parallel transporter $\Gamma[q]$ along some curve $q$ can be recovered from $\omega_a$ by path-ordered exponentiation (de Azcárraga and Izquierdo, 1995; Göckeler and Schücker, 1987)

$$\Gamma[q] = \mathrm{P}\exp\int_q \omega. \tag{B.5.5}$$

This is morally the same prescription one uses to recover a Lie group by exponentiation of its algebra.

By substituting into the gauge transformation law for the parallel transporter and keeping only first order terms, it is possible to show that the gauge transformation law for the connexion is

$$\omega' = \gamma\omega\gamma^{-1} + \gamma(\mathrm{d}\gamma^{-1}). \tag{B.5.6}$$

It is a theorem that a principal bundle $P(G, M)$ with paracompact base admits infinitely many connexions (de Azcárraga and Izquierdo, 1995). Connexions form an affine space; therefore, if $\omega_1$ and $\omega_2$ are connexion, then $\omega_1 + \omega_2$ is *not* a connexion, and, in general $\lambda_1\omega_1 + \lambda_2\omega_2$ is only a connexion when $\lambda_1 = 1 - \lambda_2$.

**B.5.2. Field strength and curvature. Cartan structural equations.** For almost all connexions, the parallel transporter $\Gamma[q]$ depends not just on the end-point of $q$ but also on the details of the path between them. Therefore, parallel transport along two different paths between the same end-points will not usually give the same result. As a special case, for example, it may happen that $\Gamma[q]$ is not the identity even for a closed path.

[3] $\mathfrak{g}$ is the Lie algebra of the group $G$.

The connexion itself can be used the measure the obstruction to triviality of parallel transport around closed paths. We define the curvature $\Omega$ of a connexion $\omega$ to be

$$\Omega = \mathrm{d}\omega + \frac{1}{2}[\omega, \omega] = \mathrm{d}\omega + \omega \wedge \omega. \tag{B.5.7}$$

This is called the Cartan structural equation, or sometimes Cartan's second structural equation. The operator $\mathcal{D}$ defined by

$$\mathcal{D}A = \mathrm{d}A + \omega \wedge A \tag{B.5.8}$$

is called the (exterior) gauge covariant derivative; in this language, Cartan's structural equation is written $\Omega = \mathcal{D}\omega$. $\Omega$ satisfies the Bianchi identity, $\mathcal{D}\Omega = 0$.

**Example.** Let $P(G, M)$ be a principal bundle, where $M$ is spacetime and $G$ a compact gauge group with Lie algebra $\mathrm{Lie}(G) = \mathfrak{g}$ (de Azcárraga and Izquierdo, 1995). In most cases, $M$ is contractible so $P(G, M)$ is trivial. (A principal bundle over a contractible manifold is always trivial.) A Yang–Mills field is a $\mathfrak{g}$-valued connexion $A_a$ on $P(G, M)$. We define functions $g(x)$ which are elements of the group of mappings $g : M \to G$ from spacetime to the group $G$. If the generators of $G$ are labelled $t_i$, then

$$g(x) = \exp\left(-\zeta^i(x)t_i\right) \simeq 1 - \zeta^i(x)t_i \quad \text{(to first order in } \zeta^i\text{)}. \tag{B.5.9}$$

The connexion obeys the transformation law (B.5.6), which in the present context says

$$A'_a(x) = g(x)A_a(x)g^{-1}(x) + g(x)\partial_a g^{-1}(x) = g(A_a + \partial_a)g^{-1}. \tag{B.5.10}$$

Therefore, an infinitesimal transformation of $A_a = A_a$ can be written

$$\begin{aligned} A'_a - A_a = \delta_\zeta A_a &= \partial_a \zeta^i(x)t_i + A_a^m \zeta^n [t_m, t_n] \\ &= \partial_a \zeta^i(x)t_i + f^i_{mn} A_a^m \zeta^n t_i \end{aligned} \tag{B.5.11}$$

using the Lie algebra relation $[t_m, t_n] = f^i_{mn} t_i$, where the $f^i_{mn}$ are known as the structure constants. This can be rewritten

$$\delta_\zeta A_a = \mathcal{D}_a \zeta(x), \quad \text{where } \zeta = \zeta^i t_i. \tag{B.5.12}$$

The curvature of $A$ is written $F = \mathcal{D}A$ and called the field strength. Under a gauge transformation, the field strength changes tensorially, $F \mapsto F^g = g^F g^{-1}$, so the variation in $F$ under an infinitesimal transform is

$$\delta_\zeta F(x) = [\zeta(x), F(x)]. \tag{B.5.13}$$

In addition, as for any curvature, the field strength satisfies the Bianchi identity $\mathcal{D}F = 0$. If a metric is present, then one can write a Lagrangian for $F$,

$$\mathcal{L} = \frac{1}{g^2} \int_M \mathrm{d}v \ \mathrm{Tr}\, F \wedge *F. \tag{B.5.14}$$

This is called pure Yang–Mills theory with coupling $g$. The field equation is $\mathcal{D} * F = 0$.

## B.6. Einstein–Cartan theory

We now apply all of the preceding technology to the construction of a dynamical theory of gravity. The theory we shall construct is not quite Einstein gravity but in circumstances where the Einstein theory is applicable (vanishing torsion; no fermions) it is equivalent.

Let $M$ be a manifold with metric of signature $\mathrm{diag}(-1, 1, \ldots, 1)$, and let $\beta^\mu$ be a frame. Physics should be invariant of the choice of frame, so Einstein–Cartan gravity is the gauge theory of $GL(4)$ applied to the bundle of linear frames over spacetime. One introduces a $\mathfrak{gl}(4)$-valued connexion, conventionally denoted $\Gamma$ and subject to the familiar transformation rule[4]

$$\Gamma' = \gamma\Gamma\gamma^{-1} + \gamma \mathrm{d}\gamma^{-1} \tag{B.6.2}$$

under $GL(4)$ transformations $\gamma$. The field strength of $\Gamma$ is written $R$,

$$R = \mathcal{D}\Gamma = \mathrm{d}\Gamma + \Gamma \wedge \Gamma, \tag{B.6.3}$$

and is usually called the Riemann curvature tensor. We also define the torsion,

$$T = \mathcal{D}\beta = \mathrm{d}\beta + \Gamma \wedge \beta. \tag{B.6.4}$$

It is for this reason that we picked the bundle of linear frames, rather than work with vector fields: one can apply the exterior calculus to forms. The definition of torsion is sometimes called Cartan's first structural equation. $R$ and $T$ each satisfy a Bianchi identity

$$\mathcal{D}R = 0 \quad \text{and} \quad \mathcal{D}T = R \wedge \beta \tag{B.6.5}$$

A connexion is called metric if it preserves the metric under parallel transport. It is easy to show that for an infinitesimal curve with tangent vector $X$ this condition is

$$X^a \partial_a g_x(v, w) + g_x(\Gamma_a X^a v, w) + g_x(v, \Gamma_a X^a w) = 0. \tag{B.6.6}$$

[4]Historical terminology has overloaded the uses of the symbol $\Gamma$ in this context. It is important to stress that $\Gamma$ is *not* the parallel transporter in this context, but the connexion. The parallel transporter around a path $C$ is given by

$$\mathrm{P}\exp \int_C \Gamma. \tag{B.6.1}$$

We try and avoid use of the parallel transporter itself as much as possible, in order to minimise confusion.

Expanding in components shows that this is equivalent to

$$\partial_a g_{ij} + g_{kj}\Gamma^k_{ai} + g_{ik}\Gamma^k_{aj} = 0. \tag{B.6.7}$$

This is sometimes expressed by writing $\mathcal{D}g = 0$, with a suitable understanding of how the commutator is to be taken. We now restrict attention to orthonormal frames, written $e^\mu$ rather than the generic $\beta^\mu$. When the frame is orthonormal and the connexion is metric it is conventional to write the connexion as $\omega$ and call it the spin connexion. In an orthonormal frame the metric is constant, so the metric condition says

$$\omega^T\eta + \eta\omega = 0 \tag{B.6.8}$$

which is just the statement that $\omega$ is valued in $\mathfrak{so}(3,1)$.

The action is taken to be

$$S_{\mathrm{E}} = -\frac{1}{2\kappa^2}\int_M \mathrm{d}v\ R \wedge *(\beta \wedge \beta), \tag{B.6.9}$$

which was first written down by Hilbert and is properly known as the Einstein–Hilbert action. Although this started out as a gauge theory of a $\mathfrak{gl}(4)$ connexion, it is now a dynamical theory of the metric in virtue of the metric condition on $\omega$. To obtain the Einstein theory, one sets the torsion to zero. This is always possible for pure gravity, but can happen only under certain hypotheses about the matter Lagrangian if gravity is coupled to other fields. In the Einstein–Cartain theory, the torsion is fixed by the field equation for $\omega$, and can be non-zero. This actually happens (for example) in supergravity theories (Galperin et al., 2001; Weinberg, 1994).

**B.6.1. Weyl gravity and conformal compensation of Einstein gravity.** There is another way to look at the action for Einstein gravity, which we briefly describe here because similar constructions appear in string theory (Chapter 3), and also because it illuminates the passage from Einstein gravity (the theory of a second rank tensor, or Poincaré spins 0 and 2) to its weak field limit (Poincaré spin 2), which is important for the entirety of Part 2. The treatment given here follows Galperin et al. (2001).

The Einstein–Hilbert action with a cosmological term is

$$S_{\mathrm{E}} = -\frac{1}{2\kappa^2}\int \mathrm{d}v\ (R + 2\Lambda). \tag{B.6.10}$$

This is invariant under coordinate diffeomorphisms, as we have described:

$$\delta x^a = \tau^a(x) \quad \text{so} \quad \delta g_{ab} = \nabla_a\tau_b + \nabla_b\tau_a. \tag{B.6.11}$$

Off-shell, a symmetric second rank tensor like $g_{ab}$ carries Poincaré spins 2, 1, 0, 0 (Weinberg, 1994). The gauge transformation described by $\tau^a$ carries away a spin 1 and a spin 0 piece, leaving spin 2 and spin 0 in $g_{ab}$. One redefines the metric tensor via

$$g_{ab} = \phi^2 \hat{g}_{ab}, \tag{B.6.12}$$

There is now an extra gauge invariance that comes from cancelling dilatations of $\phi$ and $\hat{g}_{ab}$,

$$\delta\hat{g}_{ab} = 2a(x)\hat{g}_{ab} \quad \text{and} \quad \delta\phi = -a(x)\phi. \tag{B.6.13}$$

This is a scalar gauge transformation, known as a Weyl transformation, and therefore carries away another spin 0 piece. Therefore, the redefined field $\hat{g}_{ab}$ carries only spin 2 off-shell. This is the conformal graviton field, the part of the Einstein metric field which will become the weak field graviton. For this reason, one sometimes says that the conformal or Weyl tensor $C_{abcd}$ encodes the part of the gravitational field which corresponds to gravitational waves, or, what is the same thing, weak-field perturbations $h_{ab}$.

It might seem, in analogy with the Yang–Mills case (B.5.14) described just above, that gravity ought to be the theory of pure spin 2 in the same way that the Yang–Mills field $A_a$ (after gauge fixing) is the theory of pure spin 1. However, it can be shown that there does not exist a second-order equation of motion for $\hat{g}_{ab}$ with the correct gauge invariance,[5] and that the equations of motion for conformal gravity are in fact fourth-order. To find a proper physical gravity, one must allow the extra spin 0 degree of freedom. Note that conformally invariant gravities have occasionally been invoked to support non-standard interpretations of physics (Hoyle et al., 2000).

Casting the Einstein–Hilbert action in terms of $\hat{g}_{ab}$ and $\phi$ leads to the simple result

$$S_{\mathrm{E}} = \frac{1}{\kappa^2} \int \mathrm{d}v \, \left[ 3\phi \left( \hat{\Box} - \frac{1}{6}\hat{R} \right) \phi + \Lambda\phi^4 \right], \tag{B.6.14}$$

where $\hat{\Box}$ is the Laplacian built from $\hat{g}_{ab}$ and $\hat{R}$ is its curvature. This is invariant under Weyl transformations, and one can use the Weyl invariance to fix the gauge $\phi = 1$. This choice recovers the standard Einstein gravity; we use an exactly analogous technique to deal with the Polyakov action of the quantum relativistic string in Chapter 3. On the other hand, one can switch off the conformal graviton $\hat{g}_{ab}$, by setting $\hat{g}_{ab} = \eta_{ab}$. This gives

[5] This is known as Lovelock's theorem (Lovelock, 1971).

a flat-space scalar field action,

$$S_\phi = \frac{1}{\kappa^2} \int \mathrm{d}v \, \left(3\phi\Box\phi + \Lambda\phi^4\right), \tag{B.6.15}$$

which has a kinetic term of the wrong sign. This always arises from degrees of freedom associated with Weyl transformations, and further examples can be found in Chapter 3 and Chapter 8.

APPENDIX C

# Squeezed cosmological states and the transition to semiclassical behaviour

The outstanding issue with the calculation of the power spectrum of inflationary perturbations is the assumption that the quantum variance can somehow be related to the spectrum of large scale objects, such as anisotropies in the cosmic microwave background. This difficulty is often neglected in conventional treatments, but constitutes a highly important facet of the theory. Without some explanation, or excuse, to handle the transition from quantum to classical behaviour, the delicate quantum field theory calculations outlined in this thesis are all meaningless. The fundamental assumption is the Ansatz

$$\Delta^2_{\hat{\phi}}(k) = \Delta^2_{\phi}(k) \tag{C.0.16}$$

for the quantum-to-classical transition, where the quantity on the left is the coincident two-point expectation for the quantum scalar field $\hat{\phi}$, and the quantity on the right is the stochastic power spectrum for the classical cosmic scalar field $\phi$. Remarkably this can be put on a rigorous footing, and most of the details were worked out by Polarski, Starobinsky and collaborators during the mid-nineties (Kiefer, Lesgourgues, Polarski, and Starobinsky, 1998a; Kiefer, Polarski, and Starobinsky, 1998b; Lesgourgues, Polarski, and Starobinsky, 1997; Polarski and Starobinsky, 1996).

## C.1. Squeezed states and expanding universes

The difference between quantum and classical physically lies (largely) in the uncertainty relation, that for any two non-commuting operators $A$ and $B$, where $[A, B] = \mathrm{i}C$,

$$\sigma^2_A \sigma^2_B \geqslant \frac{\langle C \rangle^2}{4}. \tag{C.1.1}$$

This is an alternative (and perhaps more familiar) characterization of the the twisting operation described in Chapter 2 which deforms a commutative classical observable algebra into its equivalent quantum representation. In the classical case, $A$ and $B$ commute, so $C = 0$ and the uncertainty relation places no restriction on our knowledge of the observables

$A$ and $B$. The origin of the semi-classical behaviour lies in the expansion of the universe, which 'squeezes' (Merzbacher, 1998) one or other of $\sigma_A^2$ or $\sigma_B^2$ so that $\sigma_{A,B}^2 \to 0$ for one operator, and for the other $\sigma_{B,A}^2 \to \infty$. Equivalently, we can take $A$ and $B$ to commute.[1] Even though its conjugate then becomes maximally uncertain, if we do not observe it then that is of no consequence.

We will perform the analysis for a quantum field propagating over a fixed de Sitter background, rather than taking into account back reaction of metric perturbations, since that is simpler. The action for a quantum field $\phi$ on de Sitter space can be written in terms of a new variable $\Upsilon = a\phi$,

$$I = \int \mathrm{d}^4x \left( \frac{1}{2}(\Upsilon')^2 - \Upsilon'\Upsilon\frac{a'}{a} + \Upsilon^2\left(\frac{a'}{a}\right)^2 - \frac{1}{2}(\nabla\Upsilon)^2 \right). \tag{C.1.2}$$

The Hamiltonian is

$$H = \int \mathrm{d}^3x \left( \frac{1}{2}\Psi^2 + \Psi\Upsilon\frac{a'}{a} + \frac{1}{2}(\nabla\Upsilon)^2 \right), \tag{C.1.3}$$

where the canonical momentum $\Psi$ is given by

$$\Psi = \frac{\delta I}{\delta\Upsilon'} = \frac{\partial\mathcal{L}}{\partial\Upsilon'} = \Upsilon' - \Upsilon\left(\frac{a'}{a}\right),$$

and the field equation for $\Upsilon$ is obtained from the Gaussian kernel of the action,

$$\Upsilon'' - \triangle\Upsilon - \frac{a''}{a}\Upsilon.$$

As outlined in Chapter 2, a general quantum field $\Upsilon$ is constructed out of elements of some quantum Hilbert space, which means that $\Upsilon$ can be written in terms of a set of basis eigenfunctions $\Upsilon_{\mathbf{k}}$. Therefore,

$$\Upsilon(\mathbf{x},\tau) = \sum_{\mathbf{k}} \left( \Upsilon_{\mathbf{k}}(\tau)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} + \Upsilon_{\mathbf{k}}^{\dagger}(\tau)\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \right)$$

where the modes $\Upsilon_{\mathbf{k}}$ are given by

$$\Upsilon_{\mathbf{k}} = a_{\mathbf{k}}\frac{\chi_{\mathbf{k}}(\tau)}{\sqrt{2k}},$$

and we restrict the summation to one half of $\mathbf{k}$-space to make the summation unambiguous, conventionally defined by $k_0 > 0$, but in fact the precise specification is immaterial.

[1] There is another way of seeing this. If the expectation number of particles is large, then we can reorder the creation and annihilation operators arbitrarily, since this only introduces differences of order 1, and $N - 1 \approx N$ for $N \gg 1$.

We will also need the quantites $\Psi_{\mathbf{k}}$, which are the Fourier transform components of $\Psi$,

$$\Psi(\mathbf{x},\tau) = \sum_{\mathbf{k}} \left( \Psi_{\mathbf{k}}(\tau) \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} + \Psi^{\dagger}_{\mathbf{k}}(\tau) \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \right). \tag{C.1.4}$$

These modes can be obtained by substituting the mode-decomposition form of $\Upsilon$ into the equation $\Psi = \Upsilon' - \Upsilon(a'/a)$ for the canonical momentum. Substituting the Fourier forms of $\Upsilon$ and $\Psi$ into the Hamiltonian, we obtain[2]

$$\begin{aligned} \int \mathrm{d}^3x \Bigg( & \frac{1}{2} \sum_{\mathbf{k}} \sum_{\mathbf{p}} \left\{ \Psi_{\mathbf{k}} \Psi^{\dagger}_{\mathbf{p}} \mathrm{e}^{\mathrm{i}\mathbf{x}\cdot(\mathbf{k}-\mathbf{p})} + \Psi^{\dagger}_{\mathbf{k}} \Psi_{\mathbf{p}} \mathrm{e}^{\mathrm{i}\mathbf{x}\cdot(\mathbf{p}-\mathbf{x})} \right\} \\ & + \frac{a'}{a} \sum_{\mathbf{k}} \sum_{\mathbf{p}} \left\{ \Psi_{\mathbf{k}} \Upsilon^{\dagger}_{\mathbf{p}} \mathrm{e}^{\mathrm{i}\mathbf{x}\cdot(\mathbf{k}-\mathbf{p})} + \Upsilon_{\mathbf{p}} \Psi^{\dagger}_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{x}\cdot(\mathbf{p}-\mathbf{k})} \right\} \\ & + \frac{1}{2} \sum_{\mathbf{k}} \sum_{\mathbf{p}} \left\{ (\mathrm{i}\mathbf{k}\Upsilon_{\mathbf{k}}) \cdot (-\mathrm{i}\mathbf{p}\Upsilon^{\dagger}_{\mathbf{p}}) \mathrm{e}^{\mathrm{i}\mathbf{x}\cdot(\mathbf{k}-\mathbf{p})} + (-\mathrm{i}\mathbf{k}\Upsilon^{\dagger}_{\mathbf{k}}) \cdot (\mathrm{i}\mathbf{p}\Upsilon_{\mathbf{p}}) \mathrm{e}^{\mathrm{i}\mathbf{x}\cdot(\mathbf{p}-\mathbf{k})} \right\} \Bigg). \end{aligned} \tag{C.1.7}$$

Summing over $\delta$-functions and collecting terms gives[3]

$$H = L^3 \sum_{\mathbf{k}} \left( \Psi_{\mathbf{k}} \Psi^{\dagger}_{\mathbf{k}} + k^2 \Upsilon_{\mathbf{k}} \Upsilon^{\dagger}_{\mathbf{k}} + \frac{a'}{a} \left( \Psi_{\mathbf{k}} \Upsilon^{\dagger}_{\mathbf{k}} + \Upsilon_{\mathbf{k}} \Psi^{\dagger}_{\mathbf{k}} \right) \right). \tag{C.1.8}$$

Now we are in a position to make progress. One can define an operator $b_{\mathbf{k}}(\tau)$ by

$$b_{\mathbf{k}}(\tau) = \frac{1}{\sqrt{2}} \left( \sqrt{k} \Upsilon_{\mathbf{k}}(\tau) + \frac{\mathrm{i}}{\sqrt{k}} \Psi_{\mathbf{k}}(\tau) \right). \tag{C.1.9}$$

At once, this looks like an annihilation operator, but it cannot annihilate genuine $\Upsilon$ particles because it does not obey the reality condition $b_{\mathbf{k}} = b^{\dagger}_{-\mathbf{k}}$.[4] Therefore, the field $b(\mathbf{x},\tau)$

---

[2]The term proportional to $a'/a$ has been reordered. This is a necessary and consistent procedure, because otherwise the reality condition is not obeyed. The naïve idea, that the ordering should be $\Psi_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{p}} + \Psi^{\dagger}_{\mathbf{k}}\Upsilon_{\mathbf{p}}$, does not work, because

$$(\Psi_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{p}} + \Psi^{\dagger}_{\mathbf{k}}\Upsilon_{\mathbf{p}})^{\dagger} = \Upsilon_{\mathbf{p}}\Psi^{\dagger}_{\mathbf{k}} + \Upsilon^{\dagger}_{\mathbf{p}}\Psi_{\mathbf{k}} \tag{C.1.5}$$

which is misordered to obey any reality condition, whereas

$$(\Psi_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{p}} + \Upsilon_{\mathbf{p}}\Psi^{\dagger}_{\mathbf{k}})^{\dagger} = \Upsilon_{\mathbf{p}}\Psi^{\dagger}_{\mathbf{k}} + \Psi_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{p}}, \tag{C.1.6}$$

which is correct.

[3]This formula should be compared with, for example, Eq. (3) of Polarski and Starobinsky (1996). The difference of a factor of 2 arises because we are summing over only half of k-space.

[4]To see this explicitly, note that $b^{\dagger}_{\mathbf{k}}$ is given by

$$b^{\dagger}_{\mathbf{k}} = \frac{1}{\sqrt{2}} \left( \sqrt{k} \Upsilon^{\dagger}_{\mathbf{k}} - \frac{\mathrm{i}}{\sqrt{k}} \Upsilon^{\dagger}_{\mathbf{k}} \right), \tag{C.1.10}$$

and in consequence $b^{\dagger}_{-\mathbf{k}}$ is just

$$b^{\dagger}_{-\mathbf{k}} = \frac{1}{\sqrt{2}} \left( \sqrt{k} \Upsilon_{\mathbf{k}} - \frac{\mathrm{i}}{\sqrt{k}} \Psi_{\mathbf{k}} \right). \tag{C.1.11}$$

given by the analytic continuation of $b_{\mathbf{k}}$ to negative $\mathbf{k}$

$$b(\mathbf{x},\tau) = \sum_{\mathbf{k}} b_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \quad (\mathbf{k} \text{ unrestricted}) \tag{C.1.12}$$

is *not real*. Instead, $b_{\mathbf{k}}$ annihilates superpositions of $\Upsilon$ particles, and is called a quasi-particle operator; the quantities it annihilates are correspondingly called quasi-particles. Its eigenstates will correspond to the squeezed states described above (Merzbacher, 1998). In terms of $b_{\mathbf{k}}$, we have

$$\Upsilon_{\mathbf{k}}(\tau) = \frac{b_{\mathbf{k}}(\tau) + b^{\dagger}_{-\mathbf{k}}(\tau)}{\sqrt{2k}}; \quad \text{and} \quad \Psi_{\mathbf{k}} = -\mathrm{i}\sqrt{\frac{k}{2}}\left(b_{\mathbf{k}}(\tau) - b^{\dagger}_{-\mathbf{k}}(\tau)\right). \tag{C.1.13}$$

This form for $b_{\mathbf{k}}$ is chosen to make the Hamiltonian come out nicely. The components we need for the explicitly $a$-independent part are $\Psi_{\mathbf{k}}\Psi^{\dagger}_{\mathbf{k}}$,

$$\Psi_{\mathbf{k}}\Psi^{\dagger}_{\mathbf{k}} = (-\mathrm{i})(\mathrm{i})\frac{k}{2}(b_{\mathbf{k}} - b^{\dagger}_{-\mathbf{k}})(b^{\dagger}_{\mathbf{k}} - b_{-\mathbf{k}}) = \frac{k}{2}(b_{\mathbf{k}}b^{\dagger}_{\mathbf{k}} - b_{\mathbf{k}}b_{-\mathbf{k}} - b^{\dagger}_{-\mathbf{k}}b^{\dagger}_{\mathbf{k}} + b^{\dagger}_{-\mathbf{k}}b_{-\mathbf{k}}) \tag{C.1.14}$$

and $\Upsilon_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{k}}$,

$$\Upsilon_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{k}} = \frac{(b_{\mathbf{k}} + b^{\dagger}_{-\mathbf{k}})(b^{\dagger}_{\mathbf{k}} + b_{-\mathbf{k}})}{2k} = \frac{b_{\mathbf{k}}b^{\dagger}_{\mathbf{k}} + b_{\mathbf{k}}b_{-\mathbf{k}} + b^{\dagger}_{-\mathbf{k}}b^{\dagger}_{\mathbf{k}} + b^{\dagger}_{-\mathbf{k}}b_{-\mathbf{k}}}{2k}. \tag{C.1.15}$$

The quantity multiplying $a'/a$ is $\Psi_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{k}}$, and its complex conjugate. This is

$$\Upsilon_{\mathbf{k}}\Psi^{\dagger}_{\mathbf{k}} = \frac{\mathrm{i}}{\sqrt{2k}}\sqrt{\frac{k}{2}}(b_{\mathbf{k}} + b^{\dagger}_{-\mathbf{k}})(b^{\dagger}_{\mathbf{k}} - b_{-\mathbf{k}}) = \frac{\mathrm{i}}{2}(b_{\mathbf{k}}b^{\dagger}_{\mathbf{k}} - b_{\mathbf{k}}b_{-\mathbf{k}} + b^{\dagger}_{-\mathbf{k}}b^{\dagger}_{\mathbf{k}} - b^{\dagger}_{-\mathbf{k}}b_{-\mathbf{k}}) \tag{C.1.16}$$

$$\Psi_{\mathbf{k}}\Upsilon^{\dagger}_{\mathbf{k}} = (-\mathrm{i})\sqrt{\frac{k}{2}}\frac{1}{\sqrt{2k}}(b_{\mathbf{k}} - b^{\dagger}_{-\mathbf{k}})(b^{\dagger}_{\mathbf{k}} + b_{-\mathbf{k}}) = \frac{\mathrm{i}}{2}(b^{\dagger}_{-\mathbf{k}}b^{\dagger}_{\mathbf{k}} + b^{\dagger}_{-\mathbf{k}}b_{-\mathbf{k}} - b_{\mathbf{k}}b^{\dagger}_{\mathbf{k}} - b_{\mathbf{k}}b_{-\mathbf{k}}). \tag{C.1.17}$$

Substituting all this into the Hamiltonian gives

$$H = L^3\sum_{\mathbf{k}}\left[k(b_{\mathbf{k}}b^{\dagger}_{\mathbf{k}} + b^{\dagger}_{-\mathbf{k}}b_{-\mathbf{k}}) + \mathrm{i}\frac{a'}{a}(b^{\dagger}_{-\mathbf{k}}b^{\dagger}_{\mathbf{k}} - b_{\mathbf{k}}b_{-\mathbf{k}})\right]. \tag{C.1.18}$$

The $a$-independent part is what one expects for a quantum harmonic oscillator. The term proportional to $a'/a$ is what is new, and is responsible for the squeezing; importantly, we should note that it scales with $a'/a$ and so depends on the existence of an expansion or contraction.

When promoted to a canonical quantum field theory, the coordinate $\Upsilon$ and its canonical momentum $\Psi$ should obey the canonical commutation relations (Chapter 2)

$$[\Upsilon(\mathbf{x}), \Psi(\mathbf{y})] = \mathrm{i}\delta^{(3)}_{\mathrm{D}}(\mathbf{x} - \mathbf{y}), \tag{C.1.19}$$

---

This is $b^{*}_{\mathbf{k}}$, not $b_{\mathbf{k}}$.

or, writing the orthogonal basis expansions explicitly,

$$\sum_{\mathbf{k}}\sum_{\mathbf{p}}\Big([\Upsilon_{\mathbf{k}},\Psi_{\mathbf{p}}]e^{i\mathbf{x}\cdot\mathbf{k}}e^{i\mathbf{y}\cdot\mathbf{p}}+[\Upsilon^\dagger_{\mathbf{k}},\Psi_{\mathbf{p}}]e^{i\mathbf{x}\cdot\mathbf{k}}e^{-i\mathbf{y}\cdot\mathbf{p}}+[\Upsilon_{\mathbf{k}},\Psi^\dagger_{\mathbf{p}}]e^{-i\mathbf{x}\cdot\mathbf{k}}e^{i\mathbf{y}\cdot\mathbf{p}}+ [\Upsilon^\dagger_{\mathbf{k}},\Psi^\dagger_{\mathbf{p}}]e^{-i\mathbf{x}\cdot\mathbf{k}}e^{-i\mathbf{y}\cdot\mathbf{p}}\Big)=i\delta^{(3)}_{\mathrm{D}}(\mathbf{x}-\mathbf{y}). \tag{C.1.20}$$

Multiplying on the left by $e^{-i\mathbf{k}'\cdot\mathbf{x}}$ and integrating over $\mathbf{x}$ gives

$$L^3\sum_{\mathbf{p}}\Big([\Upsilon_{\mathbf{k}'},\Psi_{\mathbf{p}}]e^{i\mathbf{p}\cdot\mathbf{y}}+[\Upsilon_{\mathbf{k}'},\Psi^\dagger_{\mathbf{p}}]e^{-i\mathbf{p}\cdot\mathbf{y}}\Big)=ie^{-i\mathbf{k}'\cdot\mathbf{y}}, \tag{C.1.21}$$

and subsequently multiplying this expression on the right by $e^{i\mathbf{p}'\cdot\mathbf{y}}$ and integrating over $\mathbf{y}$ gives

$$L^3[\Upsilon_{\mathbf{k}'},\Psi^\dagger_{\mathbf{p}'}]=i\delta^{(3)}_{\mathrm{D}}(\mathbf{p}'-\mathbf{k}'). \tag{C.1.22}$$

The scale has changed because we did not choose the Fourier components $\Upsilon_{\mathbf{k}}$, $\Psi_{\mathbf{k}}$ to be orthonormal. When expanded in terms of our familiar creation and annihilation operators, the scale will turn out correctly. In terms of the $b_{\mathbf{k}}$, that is

$$-i\frac{L^3}{2}\sqrt{\frac{p}{k}}\Big([b_{\mathbf{k}},b_{\mathbf{p}}]-[b_{\mathbf{k}},b^\dagger_{-\mathbf{p}}]+[b^\dagger_{-\mathbf{k}},b_{\mathbf{p}}]-[b^\dagger_{-\mathbf{k}},b^\dagger_{-\mathbf{p}}]\Big)=i\delta^{(3)}_{\mathrm{D}}(\mathbf{p}-\mathbf{k}). \tag{C.1.23}$$

Since the $b_{\mathbf{k}}$ and $b^\dagger_{\mathbf{k}}$ commute amongst themselves, by interchanging terms in one of the remaining commutators we obtain

$$L^3[b_{\mathbf{k}},b^\dagger_{-\mathbf{p}}]=\delta^{(3)}_{\mathrm{D}}(\mathbf{k}-\mathbf{p}). \tag{C.1.24}$$

**C.1.1. Time evolution of $b_{\mathbf{k}}$.** The definition of the canonical momentum $\Psi$ shows that

$$\Psi_{\mathbf{k}}=\Upsilon'_{\mathbf{k}}-\frac{a'}{a}\Upsilon_{\mathbf{k}} \tag{C.1.25}$$

so substituting our explicit expressions for $\Psi_{\mathbf{k}}$ and $\Upsilon_{\mathbf{k}}$ into this, we have

$$-i\sqrt{\frac{k}{2}}(b_{\mathbf{k}}-b^\dagger_{-\mathbf{k}})=\frac{b'_{\mathbf{k}}+(b^\dagger_{-\mathbf{k}})'}{\sqrt{2k}}-\frac{a'}{a}\frac{b_{\mathbf{k}}+b^\dagger_{-\mathbf{k}}}{\sqrt{2k}}. \tag{C.1.26}$$

The left hand side is a superposition of an operator and its adjoint, and therefore the right hand side must be also. By reshuffling terms, the whole expression is equivalent to

$$-ik(b_{\mathbf{k}}-b^\dagger_{-\mathbf{k}})=\left(b'_{\mathbf{k}}-\frac{a'}{a}b^\dagger_{-\mathbf{k}}\right)+\left((b^\dagger_{-\mathbf{k}})'-\frac{a'}{a}b_{\mathbf{k}}\right)=\mathfrak{q}_{\mathbf{k}}+\mathfrak{q}^\dagger_{-\mathbf{k}}\quad\text{(say)} \tag{C.1.27}$$

where we have set

$$\mathfrak{q}_{\mathbf{k}}=b'_{\mathbf{k}}-\frac{a'}{a}b^\dagger_{-\mathbf{k}}. \tag{C.1.28}$$

The operator $\mathfrak{q}_{\mathbf{k}}$ is not of much use in its own right, but it does show which terms must be paired to give a correct relation. Equating operators and operator adjoints on each side then gives

$$b'_{\mathbf{k}} = \mathrm{i}kb_{\mathbf{k}} + \frac{a'}{a}b^{\dagger}_{-\mathbf{k}}; \quad \text{and} \quad (b^{\dagger}_{-\mathbf{k}})' = -\mathrm{i}kb^{\dagger}_{-\mathbf{k}} + \frac{a'}{a}b_{\mathbf{k}}. \tag{C.1.29}$$

Clearly $b_{\mathbf{k}}(\tau)$ is a linear function of $b_{\mathbf{k}}(\tau_0)$ and $b^{\dagger}_{\mathbf{k}}(\tau_0)$ for some fiducial reference time $\tau_0$, and $b^{\dagger}_{-\mathbf{k}}$ must be its complex conjugate with the sign of $\mathbf{k}$ reversed, so that gives

$$b_{\mathbf{k}}(\tau) = u_{\mathbf{k}}(\tau)b_{\mathbf{k}}(\tau_0) + v_{\mathbf{k}}(\tau)b^{\dagger}_{-\mathbf{k}}(\tau_0) \tag{C.1.30}$$

$$b^{\dagger}_{\mathbf{k}}(\tau) = u^{*}_{\mathbf{k}}(\tau)b^{\dagger}_{-\mathbf{k}}(\tau_0) + v^{*}_{\mathbf{k}}(\tau)b_{\mathbf{k}}(\tau_0). \tag{C.1.31}$$

One can show that this form for $b_{\mathbf{k}}$ is reasonable just by substituting into the evolution equation for $b'_{\mathbf{k}}$. This is just a Bogoliubov transformation.

Substituting this explicit representation of $b_{\mathbf{k}}$ into the quasi-particle commutator (C.1.24) gives

$$\begin{aligned} L^3\Big(&u_{\mathbf{k}}(\tau)u^{*}_{\mathbf{p}}(\tau)[b_{\mathbf{k}}(\tau_0), b^{\dagger}_{-\mathbf{p}}(\tau_0)] + u_{\mathbf{k}}(\tau)v^{*}_{\mathbf{p}}(\tau)[b_{\mathbf{k}}(\tau_0), b_{\mathbf{p}}(\tau_0)] \\ &+ v_{\mathbf{k}}(\tau)u^{*}_{\mathbf{p}}(\tau)[b^{\dagger}_{-\mathbf{k}}(\tau_0), b^{\dagger}_{-\mathbf{p}}(\tau_0)] + v_{\mathbf{k}}(\tau)v^{*}_{\mathbf{p}}(\tau)[b^{\dagger}_{-\mathbf{k}}(\tau_0), b_{\mathbf{p}}(\tau_0)]\Big) = \delta^{(3)}_{\mathrm{D}}(\mathbf{k}-\mathbf{p}), \end{aligned} \tag{C.1.32}$$

and using (C.1.24) again gives (after exchanging the terms in one commutator and removing terms that commute amongst themselves)

$$|u_{\mathbf{k}}(\tau)| - |v_{\mathbf{k}}(\tau)| = 1. \tag{C.1.33}$$

This should be compared with the vacuum normalization constraint we found when quantizing the scalar field on de Sitter space in an arbitrary quantum vacuum (see (4.6.79) *et seq.*). This condition is equivalent to the traditional canonical Wronskian normalization, or Hermiticity of the particle propagator plus jump and continuity conditions in the path integral formulation.

Taking all of this into account, the normalization or Wronskian relation (C.1.33) (or (4.6.79)) imposes one constraint on the complex quantites $\Upsilon_{\mathbf{k}}$, so there are three free parameters. We discussed the $\alpha$-parametrization of de Sitter vacuum states in Chapter 4 (see (4.6.80)). When discussing squeezed states, it is conventional to use an alternative parametrization:

$$u_{\mathbf{k}}(\tau) = \mathrm{e}^{-\mathrm{i}\theta_{\mathbf{k}}(\tau)}\cosh r_{\mathbf{k}}(\tau); \quad \text{and} \quad v_{\mathbf{k}}(\tau) = \mathrm{e}^{\mathrm{i}\theta_{\mathbf{k}}(\tau)-2\mathrm{i}\varphi_{\mathbf{k}}(\tau)}\sinh r_{\mathbf{k}}(\tau). \tag{C.1.34}$$

The quantities appearing here are the squeezing parameter $r_{\mathbf{k}}$, the squeezing angle $\varphi_{\mathbf{k}}$ and the phase $\theta_{\mathbf{k}}$. The use of a cosh and a sinh together with two independent phases means that (C.1.33) is automatically satisfied.

**C.1.2. Squeezing.** The connexion between the our previous formalism in terms of the $\Upsilon$-particle creation and annihilation operators $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^\dagger$, and the present case, in terms of $b_{\mathbf{k}}$ and $b_{\mathbf{k}}^\dagger$, is expressed by the equalities

$$\Upsilon_{\mathbf{k}} = a_{\mathbf{k}} \frac{\chi_{\mathbf{k}}(\tau)}{\sqrt{2k}} = \frac{b_{\mathbf{k}} + b_{-\mathbf{k}}^\dagger}{\sqrt{2k}}. \tag{C.1.35}$$

and

$$\Psi_{\mathbf{k}} = \Upsilon_{\mathbf{k}}' - \frac{a'}{a}\Upsilon_{\mathbf{k}} \quad \text{implies} \quad \Psi_{\mathbf{k}} = a_{\mathbf{k}} \frac{\chi_{\mathbf{k}}(\tau)}{\sqrt{2k}} - \frac{a'}{a} a_{\mathbf{k}} \frac{\chi_{\mathbf{k}}(\tau)}{\sqrt{2k}} = -\mathrm{i}\sqrt{\frac{k}{2}}(b_{\mathbf{k}} - b_{-\mathbf{k}}^\dagger). \tag{C.1.36}$$

We can combine these to find explicit expressions for the $b_{\mathbf{k}}$ in terms of the $a_{\mathbf{k}}$. In particular, by trivial rearrangement we obtain

$$b_{\mathbf{k}} + b_{-\mathbf{k}}^\dagger = a_{\mathbf{k}}\chi_{\mathbf{k}}; \quad \text{and} \quad b_{\mathbf{k}} - b_{-\mathbf{k}}^\dagger = \frac{\mathrm{i}}{k} a_{\mathbf{k}} \left( \chi_{\mathbf{k}}' - \frac{a'}{a}\chi_{\mathbf{k}} \right). \tag{C.1.37}$$

So, in particular,

$$b_{\mathbf{k}} = \frac{1}{2} a_{\mathbf{k}} \left( \chi_{\mathbf{k}} + \frac{\mathrm{i}}{k}\chi_{\mathbf{k}}' - \frac{1}{k}\frac{a'}{a}\chi_{\mathbf{k}} \right); \quad \text{and} \quad b_{-\mathbf{k}}^\dagger = \frac{1}{2} a_{\mathbf{k}} \left( \chi_{\mathbf{k}} - \frac{\mathrm{i}}{k}\chi_{\mathbf{k}}' + \frac{\mathrm{i}}{k}\frac{a'}{a}\chi_{\mathbf{k}} \right). \tag{C.1.38}$$

Despite the odd appearance of this pair of equations (neither involves $a_{\mathbf{k}}^\dagger$), one may verify that they are consistent. (The creation operator $a_{\mathbf{k}}^\dagger$ appears in the equation for $b_{\mathbf{k}}^\dagger$, not that for $b_{-\mathbf{k}}^\dagger$, and *vice-versa*.)

We can also write $\Upsilon_{\mathbf{k}}$ in terms of $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$,

$$\begin{aligned} \Upsilon_{\mathbf{k}} &= \frac{\left(u_{\mathbf{k}} b_{\mathbf{k}}(\tau_0) + v_{\mathbf{k}} b_{-\mathbf{k}}^\dagger(\tau_0)\right) + \left(u_{\mathbf{k}}^* b_{-\mathbf{k}}^\dagger(\tau_0) + v_{\mathbf{k}}^* b_{\mathbf{k}}(\tau_0)\right)}{\sqrt{2k}} \\ &= \frac{(u_{\mathbf{k}} + v_{\mathbf{k}}^*) b_{\mathbf{k}}(\tau_0) + (u_{\mathbf{k}}^* + v_{\mathbf{k}}) b_{-\mathbf{k}}^\dagger(\tau_0)}{\sqrt{2k}}. \end{aligned} \tag{C.1.39}$$

We will often write $f_{\mathbf{k}} = (u_{\mathbf{k}} + v_{\mathbf{k}}^*)/\sqrt{2k}$ for brevity.

## C.2. Transition to semiclassical behaviour

In the limit $r_{\mathbf{k}} \to \infty$, we get $|u_{\mathbf{k}}| \approx |v_{\mathbf{k}}|$ and in physical units one can consider the right-hand side of (C.1.33) to tend to zero. So this limit corresponds to the formal classical limit $\hbar \to 0$, and in the limit of large squeezing one expects classical behaviour to pertain.

The normalization relation for the $\chi_{\mathbf{k}}$ is

$$\chi_{\mathbf{k}}\chi'_{\mathbf{k}} - \chi_{\mathbf{k}}(\chi^*_{\mathbf{k}})' = -\frac{2ik}{L^3},$$

and so since $\Upsilon_{\mathbf{k}} = a_{\mathbf{k}}\chi_{\mathbf{k}}/\sqrt{2k}$, multiplying by $a^\dagger_{\mathbf{k}}a_{\mathbf{k}}/2k$ gives

$$\Upsilon^\dagger_{\mathbf{k}}\Upsilon'_{\mathbf{k}} - \Upsilon_{\mathbf{k}}(\Upsilon^\dagger_{\mathbf{k}})' = -\frac{\mathrm{i}}{L^3}a^\dagger_{\mathbf{k}}a^\dagger_{\mathbf{k}} \tag{C.2.1}$$

and we have normal ordered this expression. Substituting our expression for $\Upsilon_{\mathbf{k}}$ in terms of $f_{\mathbf{k}}$ and $f^*_{\mathbf{k}}$ gives

$$(b^\dagger_{\mathbf{k}}b_{\mathbf{k}} - b^\dagger_{-\mathbf{k}}b_{-\mathbf{k}})(f^*_{\mathbf{k}}f'_{\mathbf{k}} - f_{\mathbf{k}}(f^*_{\mathbf{k}})') = -\frac{\mathrm{i}}{L^3}a^\dagger_{\mathbf{k}}a_{\mathbf{k}}. \tag{C.2.2}$$

For brevity of notation, we have dropped the argument indicating that the $b_{\mathbf{k}}$ are to be evaluated at time $\tau_0$; for this rest of this section, this convention is understood except where explicitly indicated. All the time dependence in this expression is carried by $f_{\mathbf{k}}$ and $f^*_{\mathbf{k}}$.

In the limit of large squeezing or $\hbar \to 0$, the right hand side of this expression goes to zero, so we conclude

$$f^*_{\mathbf{k}}f'_{\mathbf{k}} - f_{\mathbf{k}}(f^*_{\mathbf{k}})' = 0 \tag{C.2.3}$$

which can be solved to show that $f_{\mathbf{k}} = c_{\mathbf{k}}f^*_{\mathbf{k}}$. Here, $c_{\mathbf{k}}$ is a constant depending only on $\mathbf{k}$, but not on $\tau$. Therefore, we can make $f_{\mathbf{k}}$ *real* by a time-independent phase rotation:

$$f_{\mathbf{k}} \mapsto f_{\mathbf{k}}\mathrm{e}^{-\frac{\mathrm{i}}{2}\arg c_{\mathbf{k}}}. \tag{C.2.4}$$

Then we have

$$\Upsilon_{\mathbf{k}} = f_{\mathbf{k}}(\tau)(b_{\mathbf{k}} + b^\dagger_{-}\mathbf{k}) = a_{\mathbf{k}}\frac{\chi_{\mathbf{k}}(\tau)}{\sqrt{2k}}. \tag{C.2.5}$$

Define a quantum state $|\Omega^b\rangle$ by the rule that it is annihilated by all the $b_{\mathbf{k}}$,

$$b_{\mathbf{k}}|\Omega^b\rangle = 0 \quad \text{for all } \mathbf{k}. \tag{C.2.6}$$

This is the $b$-vacuum at time $\tau_0$, and in general will not coincide with the no-particle vacuum $\Omega$. In fact, $|\Omega^b\rangle$ corresponds to a Gaussian state, and time evolution preserves its Gaussianity (Merzbacher, 1998; Polarski and Starobinsky, 1996).[5]

[5]Note that in the Heisenberg picture, the state $|\Omega^b\rangle$ itself does not change in time; instead, the operators carry time dependence. But they change in time in such a way to preserve Gaussianity of $|\Omega^b\rangle$.

Let $G(\mathbf{x}, \tau)$ be some general operator, perhaps corresponding to some observable. Then its Fourier components may be expanded as a power series in the fundamental observable $\Upsilon_{\mathbf{k}}$,

$$G_{\mathbf{k}} = \sum_{m=0}^{\infty} q_m \Upsilon_{\mathbf{k}}^m. \tag{C.2.7}$$

In the $b$-vacuum, the expectation value of this component is

$$\langle \Omega^b | G_{\mathbf{k}} G_{\mathbf{k}}^\dagger | \Omega^b \rangle = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} q_m q_n^* f_{\mathbf{k}}^{m+n} \langle \Omega^b | (b_{\mathbf{k}} + b_{-\mathbf{k}}^\dagger)^m (b_{\mathbf{k}}^\dagger + b_{-\mathbf{k}})^n | \Omega^b \rangle, \tag{C.2.8}$$

where we have assumed that $f_{\mathbf{k}}$ is real (that is, that the squeezing is large). From the creation–annihilation like properties of $b_{\mathbf{k}}$, this must be just the same as

$$\langle \Omega^b | G_{\mathbf{k}} G_{\mathbf{k}}^\dagger | \Omega^b \rangle = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} q_m q_n^* m! \delta_{\mathrm{D}}(m-n) f_{\mathbf{k}}^{m+n} = \sum_{m=0}^{\infty} |q_m|^2 m! f_{\mathbf{k}}^{2m}. \tag{C.2.9}$$

Of course, this is just a real number. We now intend to work *backwards* to obtain a classical expression.

Let $y \in \mathbf{C}$, and consider the function $\rho(|y|)$ defined by

$$\rho(|y|) = A \mathrm{e}^{-\alpha |y|^2}, \tag{C.2.10}$$

where $\alpha$ is some positive quantity, and $A$ is a normalization constant. We can use the general result

$$\int_0^\infty x^n \mathrm{e}^{-ax^2} \, \mathrm{d}x = \Gamma\left(\frac{n+1}{2}\right) \Big/ 2a^{(n+1)/2} \tag{C.2.11}$$

to show that, when $m = n$,

$$\int \mathrm{d}^2 y \; \rho(|y|) y^m (y^*)^n = \int r \, \mathrm{d}r \, \mathrm{d}\theta A \mathrm{e}^{-\alpha r^2} r^m = \frac{2\pi A}{2a^{(m+2)/2}} \Gamma\left(\frac{m+2}{2}\right) = m! \frac{\pi A}{\alpha^{m+1}} \quad \text{(where } m = n\text{)}. \tag{C.2.12}$$

Choosing $\alpha = |f_{\mathbf{k}}|^{-2}$ and $\pi A / \alpha = 1$ gives

$$\int \mathrm{d}^2 y \; \rho(|y|) y^m (y^*)^n = m! f_{\mathbf{k}}^{2m} \quad \text{(where } m = n\text{)}, \tag{C.2.13}$$

and when $m \neq n$, setting $s = \max\{m, n\}$ and $t = \min\{m, n\}$, we have

$$\int \mathrm{d}^2 y \; \rho(|y|) y^m (y^*)^n = \int r \, \mathrm{d}r \, \mathrm{d}\theta A \mathrm{e}^{-\alpha r^2} r^t \mathrm{e}^{\pm \mathrm{i}\theta(t-s)} = 0, \tag{C.2.14}$$

since we integrate $\theta$ round a full circle in the Argand plane. Consequently, we can rewrite our expression for the expectation of $G_{\mathbf{k}}$ as

$$\langle \Omega^b | G_{\mathbf{k}} G_{\mathbf{k}}^\dagger | \Omega^b \rangle = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} q_m q_n^* \int \mathrm{d}^2 y \; \rho(|y|) y^m (y^*)^n. \tag{C.2.15}$$

This may seem to be a retrograde step, since we have reintroduced a double summation, but this can just be written as an integral over the absolute value of $G$:

$$\langle \Omega^b | G_{\mathbf{k}} G_{\mathbf{k}}^\dagger | \Omega^b \rangle = \int \mathrm{d}^2 y \; \rho(|y|) \, |G(y)|^2. \tag{C.2.16}$$

The field $y$ is just a dummy variable bound to the integration, so we can interpret $\rho(|y|)$ as the probability distribution for $|y|$: it is Gaussian with mean zero and dispersion $|f_{\mathbf{k}}|^2$, and says nothing at all about the phase. Moreover, this holds for each $\mathbf{k}$ mode independently, so the phases of each $\mathbf{k}$ mode are uncorrelated. This is the origin of the interpretation of the power spectrum as a Gaussian stochastic random field (Liddle and Lyth, 2000; Peacock, 1999). Of course, it does not mean that the power spectrum *is* just a classical stochastic field (it remains a quantum object), but that we cannot distinguish it from such a field, given the very high level of squeezing in the universe today.

The crucial property in this was that the field $f_{\mathbf{k}}$ became real, and therefore that the quantity $\Upsilon_{\mathbf{k}}$ factorized into a time-dependent part, and an operator quantity.

APPENDIX D

# Holography, AdS/CFT and dS/CFT

This chapter fleshes out some of the descriptions of holographic physics, and the two principal conjectured holographic dualities, AdS/CFT and dS/CFT.

## D.1. The renormalization group

Before moving on, it is useful to pause here and say a little about the scale dependence of quantum field theories which was briefly discussed above. This is an important topic which has applications in almost all areas of modern physics, including particle physics, string theory and cosmology, and will reappear in Section 5.4. For example, when discussing cosmology we can interpret cosmological evolution as the change in scale dependence of a non-gravitational theory in one dimension fewer (Section D.2). This dependence of physical theories on the scale of interest is encoded in the renormalization group (Gell-Mann and Low, 1954; Stueckelberg and Peterman, 1952).[1]

Physics always depends on scale, and the exact methods and theories which are used to answer questions posed at a given scale change depend on the energies involved. For example, at scales of order a milimeter or so, one can successfully use continuum mechanics to predict the behaviour of solids and fluids. In order to be able to do this, one needs to supply only relatively little data, such as the viscosity, $\nu$, or the density, $\rho$, of the medium. However, at the scale of the atomic nucleus continuum mechanics is a bad approximation, and one must instead employ more microscopically detailed theories such as quantum chromodynamics. Although it is possible in principle to predict continuum mechanics and its parameters from quantum electrodynamics (taken together with other appropriate theories), this is not necessary in practice. Instead, one merely measures experimental values for $\nu$ and $\rho$ and dispenses altogether with the microscopic theory.

[1]The present treatment is heavily abbreviated. The renormalization group is described in any modern field theory text, such as Deligne et al. (1999); Peskin and Schroeder (1995); Weinberg (1994).

This pattern is repeated throughout physics. When moving from smaller scales to larger scales, one usually needs to remember only a finite amount of information about the pattern of behaviour on small scales. The remaining degrees of freedom, which are not needed in the large scale description, are simply averaged over. Mathematically, this means they become variables of integration and therefore disappear in the final answer. The renormalization group is a precise way of encoding this effect. There is also another significance. Because of the uncertainty principle one must deal with all scales at once in a quantum theory, so it is not clear how this decoupling of scales should happen. The renormalization group explains why decoupling of scales exists in quantum mechanics.

It is not guaranteed that one can always proceed from smaller to larger scales, forgetting most of the information available at each step and retaining only a finite part of the description in the transition to the next stage. Theories where this occurs are called *renormalizable*. In such theories all the information from smaller scales can be absorbed into a finite number of parameters which we can measure in order to obtain a good description of physics. In the example calculation of radiative corrections in Section 2.3.3 we needed to measure only the renormalized mass $m_R$ which encoded all details of smaller scale physics: it was these smaller scale details, which were present in the dimensionally regularized Feynman integral, that produced the $1/\varepsilon$ pole. On the other hand, if one cannot carry forward all relevant information in a finite number of parameters, then the theory is said to be unrenormalizable. In an unrenormalizable theory one would have to measure an infinite number of parameters like $m_R$ in order to remove all divergences as $\varepsilon \to 0$.

**D.1.1. The Callan–Symanzik equation.** The fundamental object in a renormalization group picture of physics is the $\beta$-function $\beta_g(E)$, which describes how a particular coupling changes with the energy scale $E$,

$$\beta_g(E) = \frac{\partial g}{\partial \ln E}. \tag{D.1.1}$$

Eq. (D.1.1) is often known as the Callan–Szymanzik equation (Callan, 1970; Symanzik, 1970). If $\beta_g(E)$ is known in advance as a function of $g(E)$, then this constitutes a differential equation for $g$ which can be used to solve for the energy dependence. In a quite general theory it might be more convenient to adopt some generalized scale parameter other than the energy $E$, so the Callan–Symanzik equation is often written in terms of a scale $\mu$, which may be an energy, a mass, a lengthscale, or any other convenient measure of scale. The

formal solution is

$$\ln \frac{E}{M} = \int_{g_M}^{g_E} \frac{\mathrm{d}g}{\beta(g)}, \tag{D.1.2}$$

where $M$ is a reference energy scale. This formal solution is valid provided the $\beta$-function does not vanish between $M$ and $E$. Depending on the characteristics of the $\beta$-function, there are a variety of ways that the coupling can behave as one pushes the energy scale to asymptotically large values. (The following table is drawn from Weinberg (1994).)

In the first case, suppose the $\beta(g)$ remains positive definite as the coupling $g$ grows.

- Singularity at finite energy. If $\beta(g)$ grows sufficiently quickly, then

$$\int^{\infty} \frac{\mathrm{d}g}{\beta(g)} < \infty. \tag{D.1.3}$$

  Then the coupling $g$ must diverge at a finite value of energy,

$$E_\infty = \mu \exp \int_{g_\mu}^{\infty} \frac{\mathrm{d}g}{\beta(g)}. \tag{D.1.4}$$

  In this case, perturbation theory in $g$ must break down at or below the energy scale $E_\infty$.

- Divergence as $E \to \infty$. On the other hand, if $\beta(g)$ does not grow sufficiently fast, then the integral will not converge, and $g_E$ will diverge only at $E = \infty$.

On the other hand, $\beta(g)$ may change sign, which gives rise to more interesting effects.

- Fixed point at finite coupling. Now suppose that $\beta(g)$ remains positive-definite in some range $0 < g < g_\star$, but descends to zero at $g = g_\star$ and is negative for $g > g_\star$. According to the Callan–Symanzik equation, $g(E)$ will be increasing with increasing $E$ below $g = g_\star$, and *decreasing* with increasing $E$ above $g = g_\star$. Thus, $g = g_\star$ is stable: it is said to be a fixed point of the renomalization group flow..

  Thus, fixed points in the renormalization group flow are not described by minima of the $\beta$-function, as one might have expected, but instead by transits of $\beta(g)$ across the $g$-axis from positive to negative. (On the other hand, $\beta(g)$ may make a transit in the opposite direction, in which case the zero of $\beta(g)$ is a *repeller* rather than a fixed point.)

  If $\beta(g)$ has a simple zero at $g = g_\star$, then near the fixed point one expects that $\beta(g)$ has the form

$$\beta(g) \sim a(g - g_*) \tag{D.1.5}$$

for some positive coefficient $a$. Solving the Callan–Symanzik equation in a small neighbourhood around $g = g_*$ shows that

$$g_* - g_E \propto E^{-a}. \tag{D.1.6}$$

The scaling $E^{-a}$ is called the anomalous scaling dimension (or anomalous dimension) of the operator whose coupling $g$ describes. In other words, in matrix elements of expectation values of an operator like $g\phi^4$ one must remember to include not only the $E$-dependence of $\phi$ by also of $g$, so the composite operator may not scale like one's naïve expectation based on the simple engineering dimension of $\phi$.

- Asymptotic freedom. A final alternative is that $\beta(g)$ may start out negative and stay negative for all $E$, increasing sufficiently slowly to drive $g_E$ to zero as $E \to \infty$. In this case one says that asymptotically free, and perturbation theory becomes exponentially good as the energy tends to infinity.

## D.2. Holographic inflation

In Section 4.5.3 we discussed how the ambiguity surrounding the quantum vacuum at the Planck scale, if such a thing as a perturbative field-theoretic vacuum is still a meaningful concept, induces some uncertainity in the description of the power spectrum which is evolved by magnifying far-ultra violet oscillations at the inflationary epoch into astrophysically-sized perturbations today.

There may be other visible effects. The general inflationary kinematics which were described in Section 4.5 in terms of scalar field theory on an FRW background obeying the Einstein equations can be re-phrased as renormalization group flow between an ultraviolet and infra-red fixed point, in the sense of Section D.1 above (Halyo, 2002b; Kabat and Lifschytz, 2002; Larsen et al., 2002; Rattazzi and Zaffaroni, 2001). This gives an alternative, complementary portrait of the microphysical description of inflation which was sketched in Section 4.6.2. We follow the treatment of Larsen et al. (2002).

In this theory, the scale parameter in the gravitational theory is clearly the radius of the universe $a$. In the dS/CFT correspondence (Halyo, 2002a; Spradlin et al., 2001; Strominger, 2001) a scalar field $\phi$ with asymptotic value $\phi_0$ in the infinite past or infinite

future is dual to an operator $\mathcal{O}$ which perturbs the dual theory away from its fixed point,

$$\mathcal{L} = \mathcal{L}_{\mathrm{CFT}} + g\mathcal{O}, \tag{D.2.1}$$

where the coupling $g$ is $g = \kappa\phi_0$, and $\kappa$ is the gravitational coupling. The relevant Callan–Symanzik equation which describes how the coupling $g$ evolves with scale is

$$\beta = \frac{\partial g}{\partial \ln \mu} = \frac{\partial}{\partial \ln a}\kappa\phi = -\frac{2}{\kappa H}\frac{\mathrm{d}H}{\mathrm{d}\phi}. \tag{D.2.2}$$

In order for this to be actually useful, we need some way to obtain the $\beta$-function. For a scalar field of mass $m$ obeying the Klein–Gordon equation, neglecting interactions, the asymptotic scaling behaviour has the form $\phi = \phi_0 \mathrm{e}^{\lambda H_0 t}$, where $\lambda$ obeys the equation

$$\lambda^2 + 3\lambda + \frac{m^2}{H_0^2} = 0. \tag{D.2.3}$$

Here $H_0$ is the de Sitter radius in the asymptotic future $t \to \infty$, for definiteness. One may also consider the asymptotic past, where $t \to -\infty$. The solutions are of the form

$$\lambda_{\pm} = -\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{m^2}{H_0^2}}. \tag{D.2.4}$$

The anomalous scaling dimension determines the $\beta$ function according to the rule

$$\lambda = \frac{\partial \ln \phi_0}{\partial \ln a} = \frac{\partial \ln g}{\partial \ln \mu} = \frac{\beta}{g}. \tag{D.2.5}$$

This is the same equation governing scaling dimensions that we uncovered when studying the near-horizon behaviour of bulk scalar fields in the braneworld, Eq. (5.7.38). This happens because we are dealing with an AdS bulk. The interpretation of $\lambda$ as the scaling dimension follows from the usual holographic argument, since under a time translation $t \mapsto t + \Delta t$ the wavefunction transforms according to the law

$$\phi \mapsto \phi \mathrm{e}^{\lambda H_0 \Delta t}. \tag{D.2.6}$$

On the other hand, in the de Sitter space metric,

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{e}^{2H_0 t}\mathrm{d}\mathbf{x}^2, \tag{D.2.7}$$

such a shift is equivalent to a rescaling of $\mathbf{x}$,

$$\mathbf{x} \mapsto \mathrm{e}^{H_0 \delta t}, \tag{D.2.8}$$

or, more transparently, a change of scale on the spatial slices. In the renormalization group interpretation, the renormalizationg group flow is threading the spatial $\mathbf{x}$-slices, and time

translation is the renormalization group. This is a Wick rotation of the RG flow in comparison with the renormalization group interpretation of Verlinde compactification outlined in Section 5.4, which is unsurprising given that dS/CFT and AdS/CFT are essentially related via Wick rotation.

The interpretation of time translation as renormalization group flow leads to an elegant reformulation of the entire history of the universe in terms of RG flow. Since current astronomical data favour a recent onset of a $\Lambda$-dominated epoch which will, if unchecked, ultimately lead to an asymptotic de Sitter state for the universe,[2] the assumption of an inflationary epoch in the early universe, means that over its lifetime the universe has interpolated between two de Sitter phases at vastly different energy scales. In the dS/CFT correspondence, this would mean that interpolating expansion should be understood as the renormalization group flow linking a CFT at high energy, in the far ultra-violet (the infinite past) with a CFT at low energy, in the infra-red (corresponding to the infinite future). There is one underlying quantum field theory, with a future de Sitter fixed point and a past de Sitter fixed point, rather than the necessity to invoke a scalar field in the early universe and quintessence in the late universe. In fact, this abstraction goes some way to explain the universality of inflation. The increasing entropy of the universe should be interpreted as integrating in degrees of freedom as one flows from the ultra-violet to the infra-red. This idea can be made precise via the holographic $c$-function, which counts the degrees of freedom in the holographic dual. In dS/CFT, this is conjectured to be

$$c = \frac{1}{\kappa^2 H^2}. \tag{D.2.9}$$

Thus the $c$-function is just a description of the Hubble parameter in disguise.

[2]Asymptotic de Sitter states are problematic in quantum gravity, and many workers suggest that asymptotic de Sitter states should not exist in quantum gravity at all. Instead, de Sitter states would be interpreted as long-lived intermediary states or resonances which would eventually decay back to asymptotically flat states (cf. Kachru et al. (2003b)). The reason for the difficulty is that in pure S-matrix theories like string theory there must be a notion of asymptotic in- and out-states, which don't exist in de Sitter space. Intuitively, this happens because, given enough time, the de Sitter horizon eventually engulfs all perturbations, no matter how close to the observer, as described by the no-hair theorem (Section 4.7.) In other words, imagine any attempt to specify an asymptotic out-state $\langle a|$ which would appear as a label on the S-matrix, $S_{ab} = \langle a|\mathbf{S}|b\rangle$. Because of the de Sitter horizon, it does not matter where one specifies this out-state to be, it will eventually become inaccessible because it is carried over the horizon into a region of the universe where an observer sitting inside the horizon cannot see it.

Now consider how the dual theory approaches the infra-red. In general, this theory will contain a large number of operators, all of which will typically be important. However, as one gets close to the infra-red fixed point a considerable simplification occurs. The infra-red limit is $\mu \to 0$, so in terms of the anomalous dimension (D.1.6), operators with $a < 0$ (or $\lambda < 0$ in the language of the present section) dominate the infra-red flow. Such operators are called relevant. Operators with dimension zero, or positive dimension are called marginal and irrelevant, respectively.[3] In a given theory with a large number of operators and couplings some operators will be relevant and some irrelevant as one approaches the infra-red fixed point. The relevant operators are driving the deformation, whereas the irrelevant operators describe fluctuations around the theory. This corresponds to the division we erected above, between operators $\lambda_+$ which were infinite energy perturbations deforming the quantum theory, and operators $\lambda_-$ which were finite energy perturbations describing fluctuations. In order to have $\lambda_+$ positive, $m^2$ must be negative, so the field ought to be rolling down its potential, giving a tachyonic mass. This is the common situation in chaotic inflation.

One can equally well put the slow-roll parameters in holographic form. In particular (Larsen et al., 2002), the slow-roll parameter $\varepsilon$ satisfies

$$\varepsilon = \frac{1}{2}\beta^2, \tag{D.2.10}$$

and $\eta$ can be written

$$\eta = \frac{1}{2}\beta^2 - \frac{1}{\kappa}\frac{\mathrm{d}\beta}{\mathrm{d}\phi} \approx -\lambda. \tag{D.2.11}$$

Notice that these results can be used to rederive the standard slow-roll predictions for the amplitude and spectral indices of the scalar and tensor power spectra, and therefore provide an alternative source for the consistency relation (Chapter 7). Since no use is made of Einstein gravity in this derivation, there is a possible source of universality.

[3]The terminology is fixed by approach to the infra-red limit, so one in the ultra-violet operators with $\lambda = 0$ and $\lambda < 0$, $\lambda > 0$ are still called marginal, relevant and irrelevant. This is because the terminology was first applied to the study of critical phenomena, where one is interested in long-range, or long-wavelength effects, such as correlations over large domains as a crystal passes through a second-order phase transition. For this purpose, one is interested in the infra-red limit, and never the ultra-violet case.

# Bibliography


Abbott, B. *Analysis of first LIGO science data for stochastic gravitational waves*, 2003. gr-qc/0312088.

Aharony, O., Gubser, S., Maldacena, J., Ooguri, H., and Oz, Y. *Large N field theories, string theory and gravity.* Phys. Rept., **323** 183–386, 2000. hep-th/9905111.

Albeverio, S., Jost, J., Paycha, S., and Scarlatti, S. *A Mathematical Introduction to String Theory.* Cambridge University Press, Cambridge, 1997.

Albrecht, A. *Reply to "A different approach to Cosmology".* Physics Today, p. 44, 1999.

Allen, L. and Wands, D. *Cosmological perturbations through a simple bounce*, 2004. astro-ph/0404441.

Amendola, L. and Tocchini-Valentini, D. *Stationary dark energy: the present universe as a global attractor.* Phys. Rev. D, **64** 043509, 2001. astro-ph/0011243.

Andrews, G., Askey, R., and Roy, R. *Special Functions.* Cambridge University Press, Cambridge, 2001.

Antoniadis, I., Arkani-Hamed, N., Dimopoulos, S., and Dvali, G. *New dimensions at a millimeter to a fermi and superstrings at a TeV.* Phys. Lett. B, **436** 257–263, 1998. hep-ph/9804398.

Arkani-Hamed, N., Dimopoulos, S., and Dvali, G. *The hierarchy problem and new dimensions at a millimeter.* Phys. Lett. B, **429** 263–272, 1998. hep-ph/9803315.

Arkani-Hamed, N., Dimopoulos, S., Dvali, G., and Kaloper, N. *Infinitely Large New Dimensions.* Phys. Rev. Lett., **84** 586–589, 2000. hep-th/9907209.

Avis, S., Isham, C., and Storey, D. p. 3565, 1978.

Balasubramanian, V., Kraus, P., and Lawrence, A. Phys. Rev. D, 1999. 046003.

Banks, T. Nucl. Phys. B, **249** 332, 1985.

Barreiro, T. and Sen, A. *Generalized Chaplygin Gas in a modified gravity approach*, 2004. astro-ph/0408185.

Bartolo, N., Matarrese, S., and Riotto, A. *Adiabatic and Isocurvature Perturbations from Inflation: Power Spectra and Consistency Relations.* Phys. Rev. D, **64** 123504, 2001. astro-ph/0107502.

Bassett, B., Kunz, M., Parkinson, D., and Ungarelli, C. *Condensate cosmology – dark energy from dark matter.* Phys. Rev. D, **68** 043504, 2003. asro-ph/0211303.

Bean, R. and Melchiorri, A. *Current constraints on the dark energy equation of state.* Phys. Rev. D, **65** 041302, 2002. astro-ph/0110472.

Becchi, C., Rouet, A., and Stora, R. Comm. Math. Phys., **42** 127, 1975.

Bento, M., Bertolami, O., and Sen, A. *Generalized Chaplygin gas and CMBR constraints.* Phys. Rev. D, **67** 063003, 2003. astro-ph/0210468.

Bergshoeff, E., Kallosh, R., and Van Proeyen, A. *Supersymmetry in singular spaces.* JHEP, **0010** 033, 2000. hep-th/0007044.

Bergshoeff, E., Kallosh, R., and Van Proeyen, A. *Supersymmetry of RS bulk and brane.* Fortsch. Phys., **49** 625–632, 2001. hep-th/0012110.

Berkovits, N. *Multiloop Amplitudes and Vanishing Theorems using the Pure Spinor Formalism for the Superstring*, 2004. hep-th/0406055.

Binetruy, P., Deffayet, C., Ellwanger, U., and Langlois, D. *Brane cosmological evolution in a bulk with cosmological constant.* Phys. Lett. B, **477** 285, 2000a. hep-th/9910219.

Binetruy, P., Deffayet, C., and Langlois, D. *Non-conventional cosmology from a brane-universe.* Nucl. Phys. B, **565** 269, 2000b. hep-th/9905012.

Birrell, N. and Davies, P. *Quantum fields in curved space.* Cambridge University Press, Cambridge, 1982.

Biswas, A., Mukherji, S., and Pal, S. *Nonsingular cosmologies from branes.* Int. J. Mod. Phys. A, **19** 557–574, 2004.

Blau, S. and Guth, A. *Inflationary Cosmology.* In S. Hawking and W. Israel, eds., *300 Years of Gravitation*, pp. 524–597. Cambridge University Press, Cambridge, 1987.

Bond, J. and Efstathiou, G. Ap. J. Letters, **285** L45, 1984.

Bordag, M., Elizalde, E., and Kirsten, K. *Heat-kernel coefficients of the Laplace operator on the $d$-dimensional ball.* J. Math. Phys, **37** 895–916, 1996a. hep-th/9503023.

Bordag, M., Geyer, B., and Kirsten, K. *Zeta-function determinant of the Laplace operator on the d-dimensional ball.* Commun. Math. Phys., **179** 215–234, 1996b. hep-th/9505157.

Bordag, M., Goldhaber, A., van Nieuwenhuizen, P., and Vassilevich, D. *Heat kernels and zeta-function regularization for the mass of the supersymmetric kink.* Phys. Rev. D, **66** 125014, 2002. hep-th/0203066.

Boughn, S. and Crittenden, R. *A correlation of the cosmic microwave sky with large scale structure*, 2003. astro-ph/0305001.

Bousso, R. and Polchinski, J. *Quantization of Four-Form Fluxes and Dynamical Neutralization of the Cosmological Constant.* JHEP, **0006** 006, 2000. hep-th/0004134.

Bowcock, P., Charmousis, C., and Gregory, R. *General brane cosmologies and their global spacetime structure.* Class. Quant. Grav., **17** 4745–4764, 2000. hep-th/0007177.

Bowden, M., Taylor, A., Ganga, K., Ade, P., Bock, J., Cahill, G., Carlstrom, J., Church, S., Gear, W., Hinderks, J., Hu, W., Keating, B., Kovac, J., Lange, A., Leitch, E., Mallie, O., Melhuish, S., Murphy, J., Piccirillo, L., Pryke, C., Rusholme, B., O'Sullivan, C., and Thompson, K. *Scientific optimization of a ground-based CMB polarization experiment.* Mon. Not. Roy. Astron. Soc., **349** 321, 2004. astro-ph/0309610.

Boyanovsky, D., Cao, F., and de Vega, H. *Inflation from Tsunami-waves.* Nucl. Phys. B, **632** 121–154, 2002. astro-ph/0102474.

Brandenberger, R. *Trans-Planckian Physics and Inflationary Cosmology*, 2002. BROWN-HET-1323, hep-th/0210186.

Brandenberger, R. and Finelli, F. *On the Spectrum of Fluctuations in an Effective Field Theory of the Ekpyrotic Universe.* JHEP, **0111** 056, 2001. hep-th/0109004.

Breitenlohner, P. and Freedman, D. *Positive energy in anti-de Sitter backgrounds and gauged extended supergravity.* Phys. Lett. B, **115** 197, 1982a.

Breitenlohner, P. and Freedman, D. *Stability in gauged extended supergravity.* Ann. Phys., **144** 249, 1982b.

Bridgman, H., Malik, K., and Wands, D. *Cosmological perturbations in the bulk and on the brane.* Phys. Rev. D, **65** 043502, 2002. astro-ph/0107245.

Bucher, M., Moodley, K., and Turok, N. *The General Primordial Cosmic Perturbation.* Phys. Rev. D, **62** 083508, 2000. astro-ph/9904231.

Bucher, M., Moodley, K., and Turok, N. *Characterising the Primordial Cosmic Perturbations.* Phys. Rev. D, **66** 023528, 2002. astro-ph/0007360.

Callan, C. Phys. Rev. D, **2** 1541, 1970.

Carlip, S. *Quantum Gravity in 2+1 Dimensions.* Cambridge University Press, Cambridge, 1998.

Carroll, S. *Lecture Notes on General Relativity*, 1997. NSF-ITP/97-147, gr-qc/9112019.

Carroll, S. *Quintessence and the rest of the world.* Phs. Rev. Lett., **81** 3067–3070, 1998. astro-ph/9806099.

Ceresole, A. and Dall'Agata, G. *General matter coupled $\mathcal{N} = 2$, $d = 5$ gauged supergravity.* Nucl. Phys. B, **585** 143–170, 2000. hep-th/0004111.

Chamblin, A. and Gibbons, G. *Supergravity on the brane.* Phys. Rev. Lett., **84** 1090–1093, 2000.

Chamblin, A., Hawking, S., and Reall, H. *Brane-World Black Holes.* Phys. Rev. D, **61** 065007, 2000. hep-th/9909205.

Chamseddine, A. and West, P. Nucl. Phys. B, **129** 39, 1977.

Chandrasekhar, S. *The mathematical theory of black holes.* Oxford University Press, Oxford, 1983.

Chung, D. and Freese, K. *Lensed Density Perturbations in Braneworlds: An Alternative to Inflation.* Phys. Rev. D, **67** 103505, 2003. astro-ph/0202066.

Copeland, E., Gray, J., and Lukas, A. *Moving Five-Branes in Low-Energy Heterotic M-theory.* Phys. Rev. D, **64** 126003, 2001. hep-th/0106285.

Copeland, E., Gray, J., Lukas, A., and Skinner, D. *Five-Dimensional Moving Brane Solutions with Four-Dimensional Limiting Behaviour.* Phys. Rev. D, **66** 124007, 2002. hep-th/0207281.

Copeland, E., Gray, J., and Saffin, P. *Gravitational instantons and internal dimensions.* JHEP, **0006** 024, 2000. hep-th/0003244.

Cordero, R. and Rojas, E. *Nucleation of ${}^4R$ brane universes*, 2003. gr-qc/0302037.

Cordero, R. and Vilenkin, A. *Stealth branes.* Phys. Rev. D, **65** 083519, 2002. hep-th/0107175.

Cowsik, R. and McClelland, J. Phys. Rev. Lett., **29** 669, 1972.

Csaki, C., Kaloper, N., and Terning, J. *Dimming supernovae without cosmic acceleration.* Phys. Rev. Lett., **88** 161302, 2002. hep-ph/0111311.

Da Rold, L. *Radiative corrections in 5D and 6D: Expanding in winding modes*, 2003. hep-th/0311063.

Danielsson, U. *A eprint on inflation and transplanckian physics.* Phys. Rev. D, **66** 023511, 2002a. hep-th/0203198.

Danielsson, U. *On the consistency of de Sitter vacuab A.* 2002b. hep-th/0210058.

de Azcárraga, J. A. and Izquierdo, J. M. *Lie groups, Lie algebras, cohomology, and some applications in physics.* Cambridge University Press, Cambridge, 1995.

de Felice, F. and Clarke, C. *Relativity on curved manifolds.* Cambridge University Press, Cambridge, 1990.

D'Eath, P. *Supersymmetric Quantum Cosmology.* Cambridge University Press, Cambridge, 1996.

Deligne, P., Etingof, P., Freed, D. S., Jeffrey, L., Kazhdan, D., Morgan, J. W., Morrison, D. R., and Witten, E., eds. *Quantum Fields and Strings: A Course for Mathematicians.* American Mathematical Society, Providence, Rhode Island, 1999. Two volumes.

D'Hoker, E. and Phong, D. *Two-Loop Superstrings I, Main Formulas.* Phys. Lett. B, **529** 241–255, 2002a. hep-th/0110247.

D'Hoker, E. and Phong, D. *Two-Loop Superstrings II, The Chiral Measure on Moduli Space.* Nucl. Phys. B, **636** 3–60, 2002b. hep-th/0110283.

D'Hoker, E. and Phong, D. *Two-Loop Superstrings III, Slice Independence and Absence of Ambiguities.* Nucl. Phys. B, **636** 61–79, 2002c. hep-th/0111016.

D'Hoker, E. and Phong, D. *Two-Loop Superstrings IV, The Cosmological Constant and Modular Forms.* Nucl. Phys. B, **639** 129–181, 2002d. hep-th/0111040.

Dirac, P. *Generalized hamiltonian dynamics.* Canad. J. Math., **2** 129–148, 1950.

Dito, G. and Sternheimer, D. *Deformation Quantization: Genesis, Developments and Metamorphoses.* In Walter de Gruyter, ed., *Proceedings of the meeting between mathematicians and theoretical physicists, Strasbourg, 2001*, vol. 1 of *IRMA Lectures in Math. Theoret. Phys.* Berlin, 2002. math.QA/0201168.

Dittrich, W. and Reuter, M. *Classical and Quantum Dynamics: From Classical Paths to Path Integrals.* Springer-Verlag, Berlin, 1992.

Donini, A. and Rigolin, S. *Anisotropic Type I string compactification, winding modes and large extra dimensions.* Nucl. Phys. B, **550** 59–76, 1999. hep-ph/9901443.

Doran, M. and Jäckel, J. *Loop corrections to scalar quintessence potentials.* Phys. Rev. D, **66** 043519, 2003. astro-ph/0203018.

Durrer, R. *Clarifying perturbations in the Ekpyrotic universe*, 2001a. hep-th/0112026.

Durrer, R. *The theory of CMB anisotropies.* J. Phys. Stud., **5** 177–215, 2001b. astro-ph/0109522.

Durrer, R. and Vernizzi, F. *Adiabatic perturbations in pre big bang models: matching conditions and scale invariance.* Phys. Rev. D, **66** 083503, 2002. hep-ph/0203275.

Dvali, G. and Tye, S. *Brane inflation.* Phys. Lett. B, **450** 72–82, 1999. hep-ph/9812483.

Efstathiou, G., Moody, S., Peacock, J., Percival, W., Baugh, C., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., Colless, M., Collins, C., Couch, W., Falton, G., De Propis, R., Driver, S., Ellis, R., Frenk, C., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Norberg, P., Peterson, B., Sutherland, W., and Taylor, K. *Evidence for a non-zero Lambda and a low matter density from a combined analysis of the 2dF Galaxy Redshift Survey and Cosmic Microwave Background Anisotropies.* Mon. Not. Roy. Astron. Soc., **330** L29, 2001. astro-ph/0109152.

Elizalde, E., Nojiri, S., Odintsov, S., and Ogushi, S. *Casimir effect in de Sitter and Anti-de sitter braneworlds.* Phys. Rev. D, **67** 063515, 2003. hep-th/0209242.

Emparan, R. and Reall, H. *A rotating black ring in five dimensions.* Phys. Rev. Lett., **88** 101101, 2002. hep-th/0110260.

Erdélyi, A. *Higher Transcendental Functions.* McGraw–Hill, New York, 1953. Volume 1.

Fadeev, L. *Introduction to quantum field theory.* In P. Deligne, P. Etingof, D. S. Freed, L. Jeffrey, D. Kazhdan, J. W. Morgan, D. R. Morrison, and E. Witten, eds., *Quantum Fields and Strings: A Course for Mathematicians.* American Mathematical Society, Providence, Rhode Island, 1999. Two volumes.

Fernández Guasti, M. and Moya-Cessa, H. *Solution of the Schrödinger equation for time-dependent 1D harmonic oscillators using the orthogonal functions invariant.* J. Phys. A., **36**, 2003.

Feynman, R. and Hibbs, A. *Quantum Mechanics and Path Integrals.* McGraw–Hill, 1965.

Figueroa-O'Farrill, J. *BUSSTEPP Lectures on Supersymmetry*, 2001. hep-th/0109072.

Flachi, A., Moss, I., and Toms, D. *Quantized bulk fermions in the Randall–Sundrum brane model.* Phys. Rev. D, **64** 105029, 2001. hep-th/0106076.

Flachi, A. and Toms, D. *Quantized bulk scalar fields in the Randall–Sundrum brane model.* Nucl. Phys. B, **610** 144–168, 2001. hep-th/0103077.

Flaherty, F., ed. *Asymptotic Behaviour of Mass and Spacetime Geometry*, vol. 202 of *Lecture Notes in Physics.* Springer–Verlag, Berlin, 1984.

Fosalba, P., Gaztanaga, E., and Castander, F. *Detection of the ISW and SZ effects from the CMB–galaxy correlation*, 2003. astro-ph/0307249.

Freund, P. *Introduction to Supersymmetry.* Cambridge University Press, Cambridge, 1988.

Frolov, A. and Kofman, L. *Gravitational waves from braneworld inflation*, 2002. hep-th/0209133.

Frolov, A. and Kofman, L. *Can Inflating Braneworlds be Stabilized?* Phys. Rev. D, **69** 044021, 2004. hep-th/0309002.

Frolov, A., Kofman, L., and Starobinsky, A. *Prospects and Problems of Tachyon Matter Cosmology.* Phys. Lett. B, **545** 8–16, 2002. hep-th/0204187.

Fuchs, J. *Affine Lie Algebras and Quantum Groups.* Cambridge University Press, Cambridge, 1992.

Fujii, Y. and Maeda, K.-I. *The Scalar-tensor Theory of Gravitation.* Cambridge University Press, Cambridge, 2003.

Galperin, A., Ivanov, E., Ogievetsky, V., and Sokatchev, E. *Harmonic Superspace.* Cambridge University Press, Cambridge, 2001.

Garcia-Bellido, J. *Inflation from branes at angles*, 2003. Talk presented at Davis Inflation Meeting 2003, astro-ph/0306195.

Garcia-Bellido, J., Rabadan, R., and Zamora, F. *Inflationary scenarios from branes at angles.* JHEP, **0201** 036, 2002. hep-th/0112147.

Garriga, J. and Mukhanov, V. *Perturbations in k-inflation.* Phys. Lett. B, **458** 219–225, 1999. hep-th/9904176.

Garriga, J. and Sasaki, M. *Brane-world creation and black holes.* Phys. Rev. D, **62** 043523, 2000. hep-th/9912118.

Gell-Mann, M. and Low, F. Phys. Rev., **95** 1300, 1954.

Gibbons, G. *Thoughts on Tachyon Cosmology.* Class. Quant. Grav., **20** S321–S346, 2003. hep-th/0301117.

Gibbons, G., Hull, C., and Warner, N. Nucl. Phys. B, p. 173, 1983.

Giddings, S., Katz, E., and Randall, L. *Linearized gravity in brane backgrounds.* JHEP, **0003** 023, 2000. hep-th/0002091.

Giudice, G., Kolb, E., Lesgourgues, J., and Riotto, A. *Transdimensional physics and inflation*, 2002. hep-ph/0207145.

Göckeler, M. and Schücker, T. *Differential geometry, gauge theories, and gravity.* Cambridge University Press, Cambridge, 1987.

Goldberger, W. and Wise, M. *Modulus stabilization with bulk fields.* Phys. Rev. D, **65** 025011, 2002. hep-th/0104170.

Gorbunov, D., Rubakov, V., and Sibiryakov, S. *Gravity waves from inflating brane or mirrors moving in $AdS_5$.* JHEP, **015** 0110, 2001. hep-th/0108017.

Gordon, C., Wands, D., Bassett, B., and Maartens, R. *Adiabatic and entropy perturbations from inflation.* Phys. Rev. D, **63** 023506, 2001. astro-ph/0009131.

Gotz, M. *Validity of the 'conservation law' in the evolution of cosmological perturbations.* Mon. Not. R. Astron. Soc., **295** 873–876, 1998.

Gratton, S., Khoury, J., Steinhardt, P., and Turok, N. *Conditions for Generating Scale-Invariant Density Perturbations.* Phys. Rev. D, **69** 103505, 2004. astro-ph/0301395.

Gray, J. *An explicit example of a moduli driven phase transition in heterotic models*, 2004. DCPT-04/23, hep-th/0406241.

Gray, J. and Copeland, E. *Gravitational instantons, extra dimensions and form fields.* JHEP, **0106** 046, 2001. hep-th/0102090.

Gray, J. and Lukas, A. *Gauge Five Brane Moduli In Four-Dimensional Heterotic Models*, 2003. SUSX-TH/02-009, hep-th/0309096.

Gray, J., Lukas, A., and Probert, G. *Gauge Five Brane Dynamics and Small Instanton Transitions in Heterotic Models.* Phys. Rev. D, **69** 126003, 2004. hep-th/0312111.

Green, M., Schwarz, J., and Witten, E. *Superstring Theory.* Cambridge University Press, Cambridge, 1987. Two volumes.

Greene, B. *String theory on Calabi–Yau manifolds*, 1997. Lectures given at TASI 1996.

Gregory, R. and Padilla, A. *Braneworld instantons.* Class. Quant. Grav., **19** 279–302, 2002. hep-th/0107108.

Gubser, S. *AdS/CFT and gravity.* Phys. Rev. D, **63** 084017, 2001. hep-th/9912001.

Guedens, R., Clancy, D., and Liddle, A. *Primordial black holes in braneworld cosmologies: Formation, cosmological evolution and evaporation.* Phys. Rev. D, **66** 043513, 2002. astro-ph/0205149.

Gupta, S., Berera, A., Heavens, A., and Matarrese, S. *Non-Gaussian Signatures in the Cosmic Background Radiation from Warm Inflation.* Phys. Rev. D, **66** 043510, 2002. astro-ph/0205152.

Guth, A. Phys. Rev. D, **23** 347, 1981.

Gutowski, J. *Uniqueness of Five-Dimensional Supersymmetric Black Holes*, 2004. hep-th/0404079.

Gutowski, J. and Reall, H. *Supersymmetric AdS5 black holes.* JHEP, **0402** 006, 2004. hep-th/0401042.

Halliwell, J. and Hawking, S. *The origin of structure in the universe.* Phys. Rev. D, **31** 1777–1791, 1985.

Halyo, E. 2002a. Hep-th/0203235.

Halyo, E. *Holographic inflation*, 2002b. hep-th/0203235.

Halyo, E. *Models of inflation on D-branes*, 2003. hep-th/0307223.

Hartle, J. and Hawking, S. *Wave function of the universe.* Phys. Rev. D, **28** 2960–2975, 1983.

Hawking, S. *The quantum state of the universe.* Nucl. Phys. B, **239** 257–276, 1984.

Hawking, S. and Ellis, G. *The large scale structure of space-time.* Cambridge University Press, Cambridge, 1973.

Hawking, S., Hertog, T., and Reall, H. *Brane new world.* Phys. Rev. D, **62** 043501, 2000. hep-th/0003052.

Hawkins, E., Maddox, S., Cole, S., Madgwick, D., Norberg, P., Peacock, J., Baldry, I., Baugh, C., Bland-Hawthorn, J., and Bridges, T. *The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the universe.* MNRAS, 2002. astro-ph/0212375.

Hawkins, R. and Lidsey, J. *Inflation on a single brane – exact solutions.* Phys. Rev. D, **63** 041301, 2001. gr-qc/0011060.

Hawkins, R. and Lidsey, J. *The Ermakov–Pinney equation in scalar field cosmologies.* Phys. Rev. D, **66** 023523, 2002. astro-ph/0112139.

Helfer, A. *Do black holes radiate?* Rept. Prog. Phys., **66** 943–1008, 2003. gr-qc/0304042.

Heusler, M. *Black Hole Uniqueness Theorems.* Cambridge University Press, Cambridge, 1996.

Higgs, P. *Integration of secondary constraints in quantized general relativity.* Phys. Rev. Lett., **1** 373–374, 1958.

Hogan, C. *Holographic Discreteness of Inflationary Perturbations.* Phys. Rev. D, **66** 023521, 2002. astro-ph/0201020.

Hollands, S. and Wald, R. *An alternative to inflation*, 2002.

Horava, P. and Witten, E. *Eleven-dimensional supergravity on a manifold with boundary.* Nucl. Phys. B, **475** 94–114, 1996a. hep-th/9603142.

Horava, P. and Witten, E. *Heterotic and Type I string dynamics from eleven dimensions.* Nucl. Phys. B, **460** 506–524, 1996b. hep-th/9510209.

Horvat, R. *Observational interactions of quintessence with ordinary matter and neutrinos.* JHEP, **0208** 031, 2002. hep-ph/0007168.

Hoyle, F., Burbidge, G., and Narlikar, J. *A Different Approach to Cosmology.* Cambridge University Press, Cambridge, 2000.

Hu, W. and Sugiyama, N. *Toward understanding CMB anisotropies and their implications.* Phys. Rev. D, **51** 2599–2630, 1995. astro-ph/9411008.

Huey, G. and Lidsey, J. *Inflation, braneworlds and quintessence.* Phys. Lett. B, **514** 217–225, 2001. astro-ph/0104006.

Huey, G. and Lidsey, J. *Inflation and braneworlds: Degeneracies and consistencies*, 2002. astro-ph/0205236.

Hwang, J. *A Rigorous Proof of the Inflationary Spectrum.* J. Korean Phys. Soc., **28** S502–S511, 1995. astro-ph/9508108.

Hwang, J. and Noh, H. *Cosmological perturbations in a generalized gravity including tachyonic condensation.* Phys. Rev. D, **66** 084009, 2002. hep-th/0206100.

Ida, D., Shiromizu, T., and Ochiai, H. *Semiclassical instability of the brane-world: Randall–Sundrum bubbles.* Phys. Rev. D, **64** 023504, 2002. hep-th/0108056.

Ishibashi, A. and Wald, R. *Dynamics in Non-Globally-Hyperbolic Static Spacetimes III: Anti-de Sitter Spacetime.* Class. Quant. Grav., **21** 2981–3014, 2004. hep-th/0402184.

Israel, W. Nuovo Cimento, p. 1, 1966.

Jeffreys, H. and Jeffreys, B. *Methods of Mathematical Physics.* Cambridge University Press, Cambridge, 1946.

Johnson, C. *D-Branes.* Cambridge University Press, Cambridge, 2003.

Jones, D. *Asymptotics of the hypergeometric function.* Math. Methods. Appl. Sci., **24** 369–389, 2001.

Kabat, D. and Lifschytz, G. *Approximations for strongly-coupled supersymmetric quantum mechanics.* Nucl. Phys. B, **571** 419–456, 2000. hep-th/9910001.

Kabat, D. and Lifschytz, G. *de Sitter entropy from conformal field theory.* JHEP, **0204** 019, 2002. Hep-th/0203083.

Kachru, S., Kallosh, R., Linde, A., Maldacena, J., McAllister, L., and Trivedi, S. *Towards inflation in string theory*, 2003a. hep-th/0308055.

Kachru, S., Kallosh, R., Linde, A., and Trivedi, S. *de Sitter vacua in string theory.* Phys. Rev. D, **68** 046005, 2003b. hep-th/0301240.

Kaloper, N. *Bent Domain Walls as Braneworlds.* Phys. Rev. D, **60** 123506, 1999. hep-th/9905210.

Kaloper, N. and Linde, A. *Inflation and Large Internal Dimensions.* Phys. Rev. D, **59** 101303, 1999. hep-th/9811141.

Kamionkowski, M., Kosowsky, A., and Stebbins, A. Phys. Rev. D, **55** 7368–7388, 1997. astro-ph/9611125.

Kane, G., Perry, M., and Zytkow, A. *A possible mechanism for generating a small positive cosmological constant*, 2003. Hep-th/0311152.

Kanno, S., Sasaki, M., and Soda, J. *Born-again braneworld.* Prog. Theor. Phys., **109** 357–369, 2003. hep-th/0210250.

Kay, B. *Quantum Fields in Curved Spacetime: Non Global Hyperbolicity and Locality.* In S. Doplicher, R. Longo, J. Roberts, and L. Zsido, eds., *Operator Algebras and Quantum Field Theory.* 1996. gr-qc/9704075.

Khoury, J., Ovrut, B., Seiberg, N., Steinhardt, P., and Turok, N. *From big crunch to big bang.* Phys. Rev. D, **65** 086007, 2002a. hep-th/0108187.

Khoury, J., Ovrut, B., Steinhardt, P., and Turok, N. *The ekpyrotic universe: Colliding branes and the origin of the hot big bang.* Phys. Rev. D, **64** 123522, 2001. hep-th/0103239.

Khoury, J., Ovrut, B., Steinhardt, P., and Turok, N. *Density Perturations in the Ekpyrotic Scenario.* Phys. Rev. D, **66** 046005, 2002b. hep-th/0109050.

Khoury, J., Steinhardt, P., and Turok, N. *Inflation versus Cyclic Predictions for Spectral Tilt.* Phys. Rev. Lett., **91** 161301, 2003. astro-ph/0302012.

Kiefer, C., Lesgourgues, J., Polarski, D., and Starobinsky, A. *The Coherence of Primordial Fluctuations Produced During Inflation.* Class. Quant. Grav., **15** L67–L72, 1998a. gr-qc/9806066.

Kiefer, C., Polarski, D., and Starobinsky, A. *Quantum-to-classical transition for fluctuations in the early Universe.* Int. J. Mod. Phys. D, **7** 455–462, 1998b. gr-qc/9802003.

Klebanov, I. and Witten, E. Nucl. Phys. B, p. 89, 1999.

Kodama, H. *Uniqueness and Stability of Higher-Dimensional Black Holes*, 2004. hep-th/0403030.

Kol, B. and Wiseman, T. *Evidence that highly non-uniform black strings have a conical waist.* **20** 3493–3504, 2003. hep-th/0304070.

Kolb, E. and Turner, M. *The Early Universe.* Perseus Books, 1999.

Kolmogorov, A. and Fomin, S. *Elements of the Theory of Functions and Functional Analysis.* Graylock Press, 1957.

Koyama, K. and Soda, J. *Birth of the brane world.* Phys. Lett. B, **483** 432–442, 2000. gr-qc/0001033.

Kudoh, H. and Wiseman, T. *Properties of Kaluza–Klein black holes.* Prog. Theor. Phys., **111** 475–507, 2004. hep-th/0310104.

Langlois, D. *Inflation, quantum fluctuations and cosmological perturbations*, 2004. Lectures delivered at the Cargese School of Physics and Cosmology, Cargese, France, August 2003, hep-th/0405053.

Langlois, D., Maartens, R., and Wands, D. *Gravitational waves from inflation on the brane.* Phys. Lett. B, **489** 259–267, 2000. hep-th/0006007.

Larsen, F., van der Schaar, J., and Leigh, R. *De Sitter Holography and the Cosmic Microwave Background.* JHEP, **0204** 047, 2002. Hep-th/0202127.

Lehners, J. and Stelle, K. $d = 5$ *M-theory radion supermultiplet dynamics.* Nucl. Phys. B, **661** 273–288, 2003. hep-th/0210228.

Lesgourgues, J., Polarski, D., and Starobinsky, A. *Quantum-to-classical Transition of Cosmological Perturbations.* Nucl. Phys. B, **497** 479–510, 1997. gr-qc/9611019.

Liddle, A. *How many cosmological parameters?*, 2004. astro-ph/0401198.

Liddle, A. and Lyth, D. Phys. Rep., pp. 1–105, 1993.

Liddle, A. and Lyth, D. *Cosmological Inflation and Large-Scale Structure.* Cambridge University Press, Cambridge, 2000.

Liddle, A. and Taylor, A. *Inflaton potential reconstruction in the braneworld scenario.* Phys. Rev. D, **65** 041301, 2002. astro-ph/0109412.

Lidsey, J., Liddle, A., Kolb, E., Copeland, E., Barreiro, T., and Abney, M. *Reconstructing the inflaton potential – an overview.* Rev. Mod. Phys, **69** 373, 1997. astro-ph/9508078.

Lidsey, J., Mulryne, J., Nunes, N., and Tavakol, R. *Oscillatory Universes in Loop Quantum Cosmology and Initial Conditions for Inflation*, 2004. gr-qc/0406042.

Lidsey, J., Wands, D., and Copeland, E. *Superstring cosmology.* Phys. Rept., **337** 343–492, 2000. hep-th/9909061.

Lovelock, D. *The Einstein tensor and its generalizations.* J. Math. Phys., **12** (3) 498–501, 1971.

Lukas, A., Ovrut, B., Stelle, K., and Waldram, D. *Heterotic M-theory in five dimensions.* Nucl. Phys. B, **552** 246–290, 1999a. hep-th/9806051.

Lukas, A., Ovrut, B., Stelle, K., and Waldram, D. *The universe as a domain wall.* Phys. Rev. D, **59** 086001, 1999b. hep-th/9803235.

Lukas, A., Ovrut, B., and Waldram, D. *Cosmological solutions of Hořava–Witten theory.* Phys. Rev. D, **60** 086001, 1999c. hep-th/9806022.

Lyth, D. *The failure of cosmological perturbation theory in the new Ekpyrotic and cyclic Ekpyrotic scenarios.* Phys. Lett. B, **526** 173–178, 2002a. hep-ph/0110007.

Lyth, D. *The primordial curvature perturbation in the Ekpyrotic Universe.* Phys. Lett. B, **524** 1–4, 2002b. hep-ph/0106153.

Lyth, D., Malik, K., and Sasaki, M. *A general proof of the conservation of the curvature perturbation*, 2004. astro-ph/0411220.

Lyth, D. and Riotto, A. *Particle Physics Models of Inflation and the Cosmological Density Perturbation.* Phys. Rept., **314** 1–146, 1999. hep-ph/9807278.

Maartens, R. *Brane-World Gravity.* Living Rev. Relativity, **7**, 2004. URL http://www.livingreviews.org/lrr-2004-7.

Maartens, R., Wands, D., Bassett, B., and Heard, I. *Chaotic inflation on the brane.* Phys. Rev. D, **62** 041301, 2000. hep-ph/9912464.

Maccio, A., Quercellini, C., Mainini, R., Amendola, L., and Bonometto, S. *N-body simulations for coupled dark energy*, 2003. astro-ph/0309671.

Maldacena, J. *The large N limit of superconformal field theories and supergravity.* Adv. Theor. Math. Phys., **2** 231–252, 1998. hep-th/9711200.

Maldacena, J. *Non-Gaussian features of primordial fluctuations in single field inflationary models.* JHEP, **0305** 013, 2003a. astro-ph/0210603.

Maldacena, J. *TASI 2003 lectures on AdS/CFT*, 2003b. hep-th/0309246.

Martel, H., Shapiro, P., and Weinberg, S. *Ap. J.* **492** 29, 1998.

Martin, J. and Brandenberger, R. *The Corley-Jacobson dispersion relation and trans-Planckian inflation.* Phys. Rev. D, **65** 103514, 2002. hep-th/0201189.

Mauro, D. *Topic in Koopman–von Neumann theory9.* Ph.D. thesis, Universita defli Studi di Trieste, 2003. quant-ph/0301172.

Melchiorri, A., Mersini, L., Odman, C., and Trodden, M. *The State of the Dark Energy Equation of State.* Phys. Rev. D, **68** 043509, 2003. astro-ph/0211522.

Melchiorri, A. and Odman, C. *Dark Energy: Is it $Q$ or Lambda?*, 2002. Talk given at the XVIIIth IAP Colloquium, 'On the Nature of Dark Energy', IAP, Paris, astro-ph/0212566.

Merzbacher, E. *Quantum Mechanics.* Wiley, third ed., 1998.

Mezincescu, L. and Townsend, P. *Stability at a local maximum in higher dimensional anti-de Sitter space and application to supergravity.* Ann. Phys., **160** 406, 1985.

Minces, P. and Rivelles, V. JHEP, p. 010, 2001.

Morse, P. and Feshbach, H. *Methods of Theoretical Physics.* McGraw–Hill, 1953. Two volumes.

Moss, I. *Boundary terms for eleven-dimensional supergravity and M-theory.* Phys. Lett. B, **577** 71–75, 2003. hep-th/0308159.

Moss, I. *Boundary terms for supergravity and heterotic M-theory*, 2004. hep-th/0403106.

Motl, L. 2003. Personal communication.

Mukhanov, V. JETP Lett., **41** 493, 1985.

Mukhanov, V. *Phys. Lett. B.* **218** 17, 1998.

Mukhanov, V., Feldman, H., and Brandenberger, R. Phys. Rep., **215** 203, 1992.

Mukohyama, S., Shiromizu, T., and Maeda, K. *Global structure of exact cosmological solutions in the brane world.* Phys. Rev. D, **62** 024028, 2000. hep-th/9912287.

Nojiri, S. and Odintsov, S. *Brane world inflation induced by quantum effects.* Phys. Lett. B, **484** 118–123, 2000. hep-th/0004097.

Nojiri, S. and Odintsov, S. *Quantum cosmology, inflationary brane-world creation and dS/CFT correspondence.* JHEP, **0112** 033, 2001. hep-th/0107134.

Nojiri, S. and Odintsov, S. *(Anti-) de Sitter black holes in higher derivative gravity and dual conformal field theories.* Phys. Rev. D, **66** 044012, 2002. hep-th/0204112.

Nojiri, S. and Odintsov, S. *Quantum effects in five-dimensional brane-world: creation of de Sitter branes and particles and stabilization of induced cosmological constant.* JCAP, **0306** 004, 2003. hep-th/0303011.

Nojiri, S., Odintsov, S., and Zerbini, S. *Quantum (in)stability of dilatonic AdS backgrounds and holographic renormalization group with gravity.* Phys. Rev. D, **62** 064006, 2002. hep-th/0001192.

Olde Daalhuis, A. *Uniform asymptotic expansions for hypergeometric functions with large parameters I, II.* 2001.

Olver, F. *Asymptotics and Special Functions.* Academic Press, New York, 1974.

Ovrut, B., Pantev, T., and Park, J. *Small Instanton Transitions in Heterotic M-theory.* JHEP, **0005** 045, 2000. hep-th/0001133.

Parkinson, D., Bassett, B., and Barrow, J. *Mapping the dark energy with varying alpha,* 2003. astro-ph/0307227.

Peacock, J. *Cosmological Physics.* Cambridge University Press, Cambridge, 1999.

Peebles, P. *Principles of Physical Cosmology.* Princeton University Press, 1993.

Penzias, A. and Wilson, R. Ap. J., **142** 419, 1965.

Percival, W., Baugh, C., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., De Propis, R., Driver, S., Efstathiou, G., Ellis, R., Frenk, C., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Moody, S., Norberg, P., Peacock, J., Peterson, B., Sutherland, W., and Taylor, K. *The 2dF Redshift Survey: The power spectrum and matter content of the universe.* Mon. Not. Roy. Astron. Soc., **327** 1297, 2001. astro-ph/0105252.

Percival, W., Sutherland, W., Peacock, J., Baugh, C., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., De Propis, R., Driver, S., Efstathiou, G., Ellis, R., Frenk, C., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Moody, S., Norberg, P., Peterson, B., and Taylor, K. *Parameter constraints for flat cosmologies from CMB and 2dFGRS power spectra.* Mon. Not. Roy. Astron. Soc., p. 1068, 2002. astro-ph/0206256.

Perez-Victoria, M. *Randall–sundrum models and the regularized AdS/CFT correspondence.* JHEP, **0105** 064, 2001. hep-th/0105048.

Perry, M. *An Instability of de Sitter Space.* In G. Gibbons, S. Hawking, and S. Siklos, eds., *The Very Early Universe: Proceedings of the Nuffield Workshop, Cambridge,* pp.

459–463. 1982.

Perry, M. In F. Flaherty, ed., *Asymptotic Behavior of Mass and Spacetime Geometry*, vol. 202 of *Lecture Notes in Physics*, pp. 31–40. Springer–Verlag, Berlin, 1984.

Perry, M. *Black Holes*, 2001. Unpublished Lecture Notes (Part III Mathematical Tripos).

Peskin, M. and Schroeder, D. *An Introduction to Quantum Field Theory.* Perseus Publishing, 1995.

Polarski, D. and Starobinsky, A. *Semiclassicality and Decoherence of Cosmological Perturbations.* Class. Quant. Grav., **13** 377–392, 1996. gr-qc/9504030.

Polchinski, J. *String Theory.* Cambridge University Press, Cambridge, 1998. Two volumes.

Polnarev, A. Sov. Astron., **29** 607, 1985.

Ramírez, E. and Liddle, A. *Inflationary slow-roll formalism and perturbations in the Randall–Sundrum Type II braneworld*, 2003. astro-ph/0309608.

Randall, L. and Sundrum, R. *An alternative to compactification.* Phys. Rev. Lett., **83** 4690–4693, 1999a. hep-th/9906064.

Randall, L. and Sundrum, R. *A large mass hierarchy from a small extra dimension.* Phys. Rev. Lett., **83** 3370–3373, 1999b. hep-ph/9905221.

Rattazzi, R. and Zaffaroni, A. *Comments on the Holographic Picture of the Randall-Sundrum Model.* JHEP, **0104** 021, 2001. hep-th/0012248.

Reall, H. *Higher dimensional black holes and supersymmetry.* Phys. Rev. D, **68** 024024, 2003. hep-th/0211290.

Riesz, F. and Sz.-Nagy, B. *Functional Analysis.* Frederick Ungar Publishing Co., New York, 1955.

Riotto, A. *Inflation and the Theory of Cosmological Perturbations*, 2002. Lectures delivered at the "ICTP Summer School on Astroparticle Physics and Cosmology", Trieste, 17 June - 5 July 2002, hep-ph/0210162.

Rivers, R. *Path integral methods in quantum field theory.* Cambridge University Press, Cambridge, 1988.

Rosu, H., Espinoza, P., and Reyes, M. *Ermakov approach for Q=0 empty FRW minisuperspace oscillators.* Nuovo Cim. B, **114** 1439–1444, 1999. gr-qc/9910070.

Rovelli, C. *Loop quantum gravity.* Liv. Rev. Rel., 1997. gr-qc/9710008.

Sahni, V. and Starobinsky, A. *The Case for a Positive Cosmological Constant.* Int. J. Mod. Phys. D, **9** 373–444, 2000. astro-ph/9904398.

Santos, M., Vernizzi, F., and Ferreira, P. *Isotropy and stability of the brane.* Phys. Rev. D, **64** 063506, 2001. hep-ph/0103112.

Sanyal, A. *Quantum mechanical formulation of quantum cosmology for brane world effective action*, 2003. gr-qc/0305042.

Sasaki, M. Prog. Theor. Phys., **76** 1036, 1986.

Schmidt, H.-J. *A new proof of Birkhoff's theorem*, 1997. gr-qc/9709071.

Schoen, R. and Yau, S.-T. *Positivity of the Total Mass of a General Space-Time.* Phys. Rev. Lett., **43** 1457–1459, 1979.

Scranton, R., Connolly, A., Nichol, R., Stebbins, A., Szapudi, A., Szapudi, I., Eisenstein, D., Afshordi, N., Budavari, T., Csabai, I., Frieman, J., Gunn, J., Johnson, D., Yoh, Y., Lupton, R., Miller, C., Sheldon, E., Sheth, R., Szalay, A., Tegmark, M., and Xu, Y. *Physical evidence for dark energy*, 2003. astro-ph/0307335.

Seahra, S. *Physics in Higher Dimensional Manifolds.* Ph.D. thesis, 2003.

Seahra, S., Sepangi, H., and Ponce de Leon, J. *Brane classical and quantum cosmology from an effective action.* Phys. Rev. D, **68** 066009, 2003. gr-qc/0303115.

Seery, D. *The Cauchy Problem and Singularities in General Relativity.* Master's thesis, University of Cambridge, 2001.

Seery, D. and Bassett, B. *Radiative constraints on brane quintessence.* JCAP, **02** 001, 2004. astro-ph/0310208.

Seery, D. and Taylor, A. *Consistency relation in braneworld inflation*, 2003. astro-ph/0309512.

Sen, A. *Tachyon Condensation on the Brane Antibrane System.* JHEP, **9808** 012, 1998. hep-th/9805170.

Sen, A. *Universality of the Tachyon Potential.* JHEP, **9912** 027, 1999. hep-th/9911116.

Sen, A. *Field Theory of Tachyon Matter.* Mod. Phys. Lett. A, **17** 1797–1804, 2002a. hep-th/0204143.

Sen, A. *Tachyon Matter.* JHEP, **0207** 065, 2002b. hep-th/0203265.

Sen, A. *Remarks on Tachyon Driven Cosmology*, 2003. Talk at Nobel Symposium on Cosmology and String Theory, hep-th/0312153.

Shiromizu, T., Maeda, K., and Sasaki, M. *The Einstein equations on the 3-brane world.* Phys. Rev. D, **62** 024012, 2000. gr-qc/9910076.

Shoemaker, D. *Detector description and performance for the first coincidence observations between LIGO and GEO.* Nucl. Instrum. Meth. A, **517** 154–179, 2004. gr-qc/0308048.

Spergel, D., Verde, L., Peiris, H., Komatsu, E., Nolta, M., Bennett, C., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., M., L., Meyer, S., Page, L., Tucker, G., Weiland, J., Wollack, E., and Wright, E. *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters.* Ap. J., 2003. astro-ph/0302209.

Spradlin, M., Strominger, A., and Volovich, A. 2001. Les Houches lectures on de Sitter space; hep-th/0110007.

Starobinsky, A. Sov. Astr. Lett., **11** 133, 1985.

Steinhardt, P. and Turok, N. *A cyclic model of the universe.* Nature, 2001. hep-th/0111030.

Steinhardt, P. and Turok, N. *The Cyclic Universe: An Informal Introduction*, 2002. astro-ph/0204479.

Stephani, H., Kramer, D., MacCallum, M., Hoenselaers, C., and Herlt, E. *Exact Solutions of Einstein's Field Equations.* Cambridge University Press, Cambridge, second ed., 2003.

Sternheimer, D. *Deformation Quantization: Twenty Years After.* In *Proceedings of the 1998 Lodz conference "Particle, Fields and Gravitation"*, vol. 453 of *AIP Conf. Proc.*, pp. 107–145. AIP Press, 1998. math.QA/9809056.

Stewart, E. and Lyth, D. *A more accurate analytic calculation of the spectrum of cosmological perturbations produced during inflation.* Phys. Lett. B, **302** 171, 1993. gr-qc/9302019.

Stewart, J. *Advanced general relativity.* Cambridge University Press, Cambridge, 1991.

Strominger, A. JHEP, 2001. 049.

Stueckelberg, E. and Peterman, A. Helv. Phys. Acta., **26** 499, 1952.

Susskind, L. *The World as a Hologram.* J. Math. Phys., **36** 6377–6396, 1995. hep-th/9409089.

Susskind, L. *The anthropic landscape of string theory*, 2003. hep-th/0302219.

Susskind, L. *Supersymmetry Breaking in the Anthropic Landscape*, 2004. hep-th/0405189.

Susskind, L., Thorlacius, L., and Uglum, J. *The stretched horizon and black hole complementarity.* Phys. Rev. D, **48** 3743–3761, 1993. hep-th/9306069.

Symanzik, K. Commun. Math. Phys., **18** 227, 1970.

Temme, N. *Large parameter cases of the Gauss hypergeometric function.* In *The Sixth International Symposium on Orthogonal Polynomials, Special Functions and their Applications.* 2001.

Thiemann, T. *Lectures on loop quantum gravity.* Lect. Notes Phys., **631** 41–135, 2003. gr-qc/0210094.

Tocchini-Velentini, D. and Amendola, L. *Stationary dark energy with a baryon-dominated era: solving the coincidence problem with a linear coupling.* Phys. Rev. D, **65** 063508, 2002. astro-ph/0108143.

Tolley, A., Turok, N., and Steinhardt, P. *Cosmological Perturbations in a Big Crunch/Big Bang Space-time.* Phys. Rev. D, **69** 106005, 2004. hep-th/0306109.

Townsend, P. *Quintessence from M-theory.* JHEP, **0111** 042, 2001. hep-th/0110072.

Turok, N., Perry, M., and Steinhardt, P. *M Theory Model of a Big Crunch/Big Bang Transition*, 2004. hep-th/0408083.

Tyutin, I. 1975. Lebedev Institute Preprint N39.

Verlinde, H. *Holography and compactification.* Nucl. Phys. B, **580** 264–274, 2000. hep-th/9906182.

Verlinde, H. *Strings/branes and cosmology*, 2003. Lectures given at the Institute for Advanced Study Programme "Prospects in Theoretical Physics – Cosmology, Particles and Strings", 2003.

Vernizzi, F. *On the conservation of second-order cosmological perturbations in a scalar field dominated universe*, 2004. astro-ph/0411463.

Vilenkin, A. and Shellard, E. *Cosmic Strings and other Topological Defects.* Cambridge University Press, Cambridge.

Visser, M. *Lorentzian Wormholes.* Springer-Verlag, Berlin and Heidelberg, 1996.

Wainwright, J. and Ellis, G., eds. *Dynamical Systems in Cosmology.* Cambridge University Press, 1997.

Wald, R. Phys. Rev. D, p. 2118, 1983.

Wald, R. *General Relativity.* Chicargo University Press, Chicargo, 1984.

Wald, R. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* Chicargo Lectures in Physics. Chicargo University Press, Chicargo, 1994.

Wands, D., Malik, K., Lyth, D., and Liddle, A. *A new approach to the evolution of cosmological perturbations on large scales.* Phys. Rev. D, **62** 043527, 2000. astro-ph/

0003278.

Watson, G. *Asymptotic expansions of hypergeometric functions.* Trans. Cambridge Phil. Soc., **22** 277–308, 1918.

Weinberg, S. *Gravitation and Cosmology.* John Wiley & Sons, Inc., New York, 1972.

Weinberg, S. Ann. N.Y. Acad. Sci., **262** 409, 1975.

Weinberg, S. Phys. Rev. Lett., **59** 2607, 1987.

Weinberg, S. *The Quantum Theory of Fields.* Cambridge University Press, Cambridge, 1994. Three volumes.

Weinstein, A. *Deformation Quantization*, 1994. URL `http://math.berkeley.edu/%7Ealanw/Bourbaki.pdf`. Notes from Seminar Bourbaki.

White, M. and Hu, W. *The Sachs-Wolfe effect.* Astron. Astrophys., **321** 8–9, 1997. `astro-ph/9609105`.

Wiltshire, D. *An introduction to quantum cosmology.* In B. Robson, N. Visvanathan, and W. Woolcock, eds., *Cosmology: the Physics of the Universe*, pp. 473–531. World Scientific, Singapore, 1996. `gr-qc/0101003`.

Witten, E. *A Simple Proof of the Positive Energy Theorem.* Commun. Math. Phys., **80** 381, 1981.

Witten, E. *Strong coupling expansion of Calabi-Yau compactification.* Nucl. Phys. B, **471** 135–158, 1996. `hep-th/9602070`.

Witten, E. Adv. Theor. Math. Phys., p. 253, 1998.

Witten, E. 1999a. URL `http://www.itp.ucsb.edu/online/susy{\_}c99/discussion/`. Remarks at ITP Santa Barbara conference "New Dimensions in Field Theory and String Theory".

Witten, E. *Perturbative quantum field theory.* In P. Deligne, P. Etingof, D. S. Freed, L. Jeffrey, D. Kazhdan, J. W. Morgan, D. R. Morrison, and E. Witten, eds., *Quantum Fields and Strings: A Course for Mathematicians.* American Mathematical Society, Providence, Rhode Island, 1999b. Two volumes.

Witten, E. *Small Instantons in String Theory.* Nucl. Phys. B, **460** 541–559, 1999c. `hep-th/9511030`.

Witten, E. *Perturbative Gauge Theory As A String Theory In Twistor Space*, 2003. `hep-th/0312171`.

Wong, Z. and Guo, D. *Special Functions.* World Scientific, Singapore, 1989.

Zaldarriaga, M. and Seljak, U. Phys. Rev. D, **55** 1830–1840, 1997. astro-ph/9609170.

Zel'dovich, Y. B. Zh. Eksp. Theor. Fiz. Pis'ma Red., **4** 174, 1966.

# Index


$(p, q)$-tensor, 361
$\beta$ decay, 101
$\beta$-function, 394
$\zeta$-function regularization, 353
$\zeta$-function, 354
$\zeta$-function regularization, 44
$n$th order residual approximant, 348
$p$-form
  closed, 368
  exact, 368
$q$-boundary, 370
$q$-chain, 370
$q$-cycle, 370
1-particle irreducible, 39
2dF galaxy redshift survey, 152

Abbott (2003), 246, 401
abundance by mass of $He^4$, 102
adiabatic vacuum, 141
ADM decomposition, 279
AdS/CFT correspondence, 38, 180, 276
affine Lie algebra, 28
affine space, 377
Aharony et al. (2000), 180, 274, 401
Albeverio et al. (1997), 24, 401
Albrecht (1999), 88, 401
Allen and Wands (2004), 179, 401
Amendola and Tocchini-Valentini (2001), 211, 401
Andrews et al. (2001), 336, 401
Anglo-Australian Observatory, 154
anisotropic stress, 105
anomalous scaling dimension, 396
anti-de Sitter space, 161
Antoniadis et al. (1998), 160, 173, 401
Arkani-Hamed et al. (1998), 160, 173, 401
Arkani-Hamed et al. (2000), 160, 401
associated bundle, 375
asymptotic freedom, 396
asymptotic series, 36
asymptotic twistor equation, 190
Avis et al. (1978), 195, 401

Balasubramanian et al. (1999), 195, 200, 401
Banks (1985), 147, 401
bare mass, 41
bare propagator, 40
Barreiro and Sen (2004), 74, 401
Bartolo et al. (2001), 255, 401
baryon number, 99
baryon to photon ratio, 99
base manifold, 373
Basset function, 231
Bassett et al. (2003), 210, 402
Bean and Melchiorri (2002), 149, 402
Becchi et al. (1975), 56, 402
Bento et al. (2003), 74, 402
Berezin integration, 67
Bergshoeff et al. (2000), 190, 402

Bergshoeff et al. (2001), 190, 402
Berkovits (2004), 44, 402
Betti number, 78, 369
Bianchi identity, 16, 169, 378, 379
Big Crunch, 117
Binétruy–Deffayet–Ellwanger–Langlois models, 160
Binetruy et al. (2000a), 160, 161, 174, 219, 264, 277, 285, 289, 402
Binetruy et al. (2000b), 160, 161, 219, 277, 285, 286, 289, 295, 302, 402
BIons, 178
Birrell and Davies (1982), 140–142, 194, 257, 402
Biswas et al. (2004), 276–278, 296, 301, 303, 308, 402
Blau and Guth (1987), 124, 245, 402
Bogoliubov transformation, 388
Bond and Efstathiou (1984), 246, 402
Bordag et al. (1996a), 328, 331, 402
Bordag et al. (1996b), 328, 331, 402
Bordag et al. (2002), 328, 402
Bose–Einstein distribution, 93
Boughn and Crittenden (2003), 210, 403
boundary conditions
  Dirichlet–Neumann, 344
  Robin, 344
Bousso and Polchinski (2000), 151, 403
Bowcock et al. (2000), 166, 186, 262, 287, 315, 403
Bowden et al. (2004), 152, 246, 403
Boyanovsky et al. (2002), 119, 403
BPS state, 176
Brandenberger and Finelli (2001), 179, 403
Brandenberger (2002), 123, 129, 403
Brans–Dicke gravity, 210
Breitenlohner and Freedman (1982a), 181, 195, 201, 287, 403
Breitenlohner and Freedman (1982b), 181, 195, 201, 287, 403
Bridgman et al. (2002), 192, 247, 403
BRST charge operator, 59
BRST quantization, 52, 56
BRST transformation, 57
Bucher et al. (2000), 154, 403
Bucher et al. (2002), 154, 403
Bunch–Davies, 141

Callan–Szymanzik equation, 394
Callan (1970), 394, 403
canonical momenta, 24
canonical quantization, 28
Carlip (1998), 24, 194, 274, 275, 291, 314, 403
Carroll (1997), 16, 404
Carroll (1998), 212, 404
Cartan structural equation, 378
Cartan's first structural equation, 379
Carter–Robinson theorem, 337
Ceresole and Dall'Agata (2000), 168, 195, 404
Chamblin and Gibbons (2000), 199, 290, 404
Chamblin et al. (2000), 191, 404
Chamseddine and West (1977), 16, 404
Chandrasekhar (1983), 176, 337, 358, 404
chaotic inflation, 119
Chaplygin gas, 74
chart, 367
Chern–Simons, 172
chiral spinors, 20
Chung and Freese (2003), 183, 404
classical singularity theorems, 189
clocks, 125
closed strings, 68
Codacci's equation, 22, 186
codifferential, 366
cohomology class, 78
cohomology classes, 369
cohomology ring, 370
comoving hypersurfaces, 126

comoving observers, 126
compactification manifold, 76
complexification, 213
configuration manifold, 26
configuration space, 24
conformal field theory, 65
conformal time, 105
conformal transformation, 65
connexion, 375
consistent truncation, 168
constraints, 49
  first class, 50
  second class, 50
coordinate realization, 28
coordinate representaton, 283
Copeland et al. (2000), 276, 404
Copeland et al. (2001), 174, 404
Copeland et al. (2002), 174, 404
Copernican principle, 87
Cordero and Rojas (2003), 276, 404
Cordero and Vilenkin (2002), 276, 404
correlation functions, 37
cosmic microwave background, 88, 112, 275
cosmic strings, 89
cosmological constant, 92
cosmological principle, 87
cotangent space, 360
Coulomb gauge, 52
Cowsik and McClelland (1972), 98, 404
Csaki et al. (2002), 210, 404
curvature
  of connexion, 378
cyclic model, 175

D'Eath (1996), 24, 132, 274, 280, 296, 301, 405
D'Hoker and Phong (2002a), 44, 405
D'Hoker and Phong (2002b), 44, 405
D'Hoker and Phong (2002c), 44, 405
D'Hoker and Phong (2002d), 44, 405
D-branes, 75, 78, 82
D-strings, 83
Danielsson (2002a), 141, 404
Danielsson (2002b), 141, 405
dark energy, 149, 209
dark radiation, 163
Da Rold (2003), 317, 404
de Rham $p$-coboundaries, 368
de Rham $p$-cocycles, 368
de Rham cohomology groups, 368
de Witt metric, 283
deceleration parameter, 91
decoherence
  of quantum fluctuations, 383
decoupling, 88, 97
decoupling temperature, 98
degeneracies, 154
degree, 359
Deligne et al. (1999), 24, 26, 28, 357, 358, 393, 405
density fluctuations, 89
density parameters, 91
deuterium bottleneck, 102
de Azcárraga and Izquierdo (1995), 28, 48, 58, 168, 172, 357, 377, 378, 405
de Felice and Clarke (1990), 357, 405
differentiable manifold, 367
differential form, 360
dilaton, 74
dimensional reduction, 211
dimensional regularization, 44
Dirac adjoint, 20
Dirac algebra, 18
Dirac bracket, 51
Dirac index, 75
Dirac inner product, 20
Dirac matrices, 18

Dirac spinor, 20
Dirac (1950), 274, 275, 405
direct product compactifications, 197
dispersion, 137
Dito and Sternheimer (2002), 29, 405
Dittrich and Reuter (1992), 305, 405
Donini and Rigolin (1999), 317, 318, 405
Doran and Jäckel (2003), 213, 216, 217, 234, 236, 240, 405
dressed propagator, 40
dS/CFT correspondence, 396
dual basis, 359
dual form, 365
Durrer and Vernizzi (2002), 179, 406
Durrer (2001a), 179, 405
Durrer (2001b), 112, 405
Dvali and Tye (1999), 159, 406

Efstathiou et al. (2001), 154, 406
Einstein gravity., 15
Einstein–de Sitter universe, 93
Einstein–Hilbert action, 162, 220
Ekpyrotic model, 175
electric field, 51
electroweak force, 48
Elizalde et al. (2003), 276, 406
Emparan and Reall (2002), 337, 406
energy density, 94
energy–momentum, 18
energy–momentum tensor, 68
entropy of the universe, 124
entropy perturbation, 127
Erdélyi (1953), 333, 336, 406
Euclidean section, 213
Euler number, 65
Euler's equation, 96
Euler–Poincaré invariant, 369
event horizon, 122
expansion into diagrams, 35
exterior algebra, 359
exterior derivative, 361
extrinsic curvature, 144

F-strings, 83
Fadeev–Popov, 52
Fadeev (1999), 24, 27, 30, 48, 51, 406
Fermi–Dirac distribution, 93
Fernández Guasti and Moya-Cessa (2003), 305, 406
Feynman and Hibbs (1965), 24, 406
Feynman diagrams, 35
Feynman graphs, 34
Feynman integrals, 35
Feynman rules, 35
Feynman series, 34
fibre over $x$, 373
field strength, 378
Figueroa-O'Farrill (2001), 18, 406
first Chern class, 168
first quantization, 23
Flachi and Toms (2001), 193, 323, 328, 330, 331, 406
Flachi et al. (2001), 193, 328, 406
Flaherty (1984), 183, 189, 406
flatness problem, 116
fluid-flow formalism, 106
foliations of spacetime, 125
Fosalba et al. (2003), 210, 406
fractional ionization, 99
frame, 361
free field theory, 34
freeze-out, 97
Freund (1988), 16, 18, 167, 169, 170, 407
Friedman–Robertson–Walker metric, 90
Friedmann equation, 91

Frolov and Kofman (2002), 192, 247, 250, 322, 407
Frolov and Kofman (2004), 169, 191, 248, 407
Frolov et al. (2002), 191, 407
Fuchs (1992), 357, 358, 407
Fujii and Maeda (2003), 339, 407
functional $\zeta$-function, 354
functional determinant, 353
future null infinity, 190

Göckeler and Schücker (1987), 18, 170, 357, 358, 377, 407
Galperin et al. (2001), 18, 24, 167, 380, 407
Garcia-Bellido et al. (2002), 119, 159, 407
Garcia-Bellido (2003), 159, 407
Garriga and Mukhanov (1999), 134, 407
Garriga and Sasaki (2000), 276, 407
gauge artefacts, 105
gauge covariant derivative, 378
gauge field theory, 47
gauge invariant
  variables, 125
gauge-fixed Polyakov action, 67
gauge-invariant description of perturbations, 125
gauge-invariant observables, 105
Gauss' equation, 22, 184
  generalized, 22
Gauss' Law, 52
Gauss–Codacci, 279
Gauss–Codacci equations, 144
Gaussian-averaged gauge, 57
Gell-Mann and Low (1954), 393, 407
generalized $\xi$-gauge, 55
generalized Gauss' equation, 22
generalized Lorentz gauge, 54
generalized Newtonian gauge, 105
generalized-$\xi$ gauge, 57
Geroch's theorem, 279
ghosts, 56
Gibbons et al. (1983), 195, 407
Gibbons–Hawking, 281
Gibbons–Hawking term, 33
Gibbons (2003), 191, 407
Giddings et al. (2000), 182, 219, 228, 231, 407
Giudice et al. (2002), 197, 247, 250, 290, 407
Goldberger and Wise (2002), 191, 408
Gorbunov et al. (2001), 200, 247, 250, 253, 265, 275, 317, 321, 322, 408
Gordon et al. (2001), 127, 408
Gotz (1998), 127, 131, 408
Grassman algebra, 359
Grassmann differentiation, 67
Grassmann integration, 67
Gratton et al. (2004), 179, 408
Gray and Copeland (2001), 276, 408
Gray and Lukas (2003), 174, 408
Gray et al. (2004), 174, 408
Gray (2004), 174, 175, 191, 408
Green et al. (1987), 24, 61, 70, 75–77, 82, 168, 273, 357, 358, 408
Green–Schwarz superstring, 76
Greene (1997), 75, 168, 178, 408
Gregory and Padilla (2002), 276, 408
Gubser (2001), 182, 286, 310, 311, 319, 408
Guedens et al. (2002), 191, 338, 408
Gupta et al. (2002), 137, 408
Guth (1981), 118, 245, 408
Gutowski and Reall (2004), 191, 337, 409
Gutowski (2004), 337, 409

Halliwell and Hawking (1985), 275, 409
Halyo (2002a), 201, 396, 409
Halyo (2002b), 119, 396, 409
Halyo (2003), 119, 409
Hamilton's equations, 25
Hamiltonian, 25

Hamiltonian form, 24
Hamiltonian vector field, 25
harmonic, 372
Hartle and Hawking (1983), 274, 409
Hausdorff space, 53
Hawking and Ellis (1973), 21, 22, 91, 149, 157, 179, 183, 184, 189, 195, 196, 249, 279, 337, 358, 409
Hawking et al. (2000), 182, 250, 276, 409
Hawking (1984), 274, 409
Hawkins and Lidsey (2001), 289, 409
Hawkins and Lidsey (2002), 305, 409
Hawkins et al. (2002), 87, 245, 409
Helfer (2003), 124, 409
Hermitian metric, 167
Hermitian operator, 139
heterotic string theory, 83
Heusler (1996), 337, 409
hierarchy problem, 182
Higgs (1958), 275, 409
Hodge decomposition theorem, 78, 372
Hodge dual, 171, 365
Hodge star, 365
Hodge–de Rham operator, 367
Hodge–de Rham theory, 76, 371
Hogan (2002), 124, 409
Hollands and Wald (2002), 123, 409
holographic $c$-function, 398
holographic projection, 180
holonomy, 167
Hopf bundle, 373
Horava and Witten (1996a), 84, 85, 170, 274, 315, 410
Horava and Witten (1996b), 84, 170, 274, 315, 410
Horvat (2002), 213, 410
Hot Big Bang model, 87
Hoyle et al. (2000), 88, 177, 381, 410
Hu and Sugiyama (1995), 112, 245, 410
Hubble length, 122
Hubble time, 122
Huey and Lidsey (2001), 203, 247, 253, 256, 275, 410
Huey and Lidsey (2002), 247, 275, 410
Hwang and Noh (2002), 130, 131, 410
Hwang (1995), 140, 202, 410

Ida et al. (2002), 276, 410
indices
  raising and lowering, 365
induced curvature, 125
induced metric, 364
inflation, 89
infra-red divergence, 43
instantons, 276
integral cohomology class, 369
integration, on $SL^p(a,b)$, 352
intrinsic curvature perturbation, 125
invariant volume form, 365
irrelevant operator, 399
Ishibashi and Wald (2004), 195, 410
isocurvature perturbation, 127
Israel junction condition, 186
Israel (1966), 162, 186, 410

Jeffreys and Jeffreys (1946), 258, 259, 410
Johnson (2003), 61, 80, 82, 84, 178, 180, 181, 315, 410
Jones (2001), 336, 410

Kähler class, 168
Kähler form, 168
Kähler potential, 167
Kabat and Lifschytz (2000), 307, 410
Kabat and Lifschytz (2002), 396, 410
Kac–Moody algebra, 28
Kachru et al. (2003a), 119, 411

Kachru et al. (2003b), 119, 158, 190, 208, 398, 411
Kaloper and Linde (1999), 161, 411
Kaloper–Linde model, 166, 197, 199, 250
Kaloper (1999), 161, 411
Kaluza–Klein fields, 220
Kaluza–Klein mechanism, 75
Kaluza–Klein modes, 80
Kamionkowski et al. (1997), 246, 411
Kane et al. (2003), 151, 207, 411
Kanno et al. (2003), 315, 411
Kay (1996), 195, 411
Kerr black hole, 337
Kerr–Newman blac hole, 337
Khoury et al. (2001), 175, 315, 411
Khoury et al. (2002a), 175, 315, 411
Khoury et al. (2002b), 175, 179, 411
Khoury et al. (2003), 179, 411
Kiefer et al. (1998a), 383, 411
Kiefer et al. (1998b), 383, 411
Klebanov and Witten (1999), 195, 200, 412
Kodama (2004), 337, 412
Kol and Wiseman (2003), 337, 412
Kolb and Turner (1999), 92, 98, 412
Kolmogorov and Fomin (1957), 251, 343, 412
Koyama and Soda (2000), 276–278, 286, 296, 299, 301, 303, 304, 412
Kronecker delta, 359
Kudoh and Wiseman (2004), 191, 337, 412

Lagrange multipliers, 49
Lagrangian
  quasi-invariant, 48
  singular, 48
Lagrangian formulation, 26
Langlois et al. (2000), 193, 203, 247, 250–253, 265, 275, 290, 315, 317, 321–323, 325, 412
Langlois (2004), 89, 126, 412
Lanzcos–Israel junction condition, 186
Laplace–de Rham operator, 367
lapse function, 279
Larsen et al. (2002), 119, 396, 399, 412
Lehners and Stelle (2003), 176, 178, 315, 316, 412
Lennard–Jones 6/12 potential, 176
Lesgourgues et al. (1997), 383, 412
Levi–Civita tensor, 365
Liddle and Lyth (1993), 89, 106, 136–138, 412
Liddle and Lyth (2000), 88, 89, 106, 109, 113, 115, 127, 138, 142, 153, 179, 246, 392, 412
Liddle and Taylor (2002), 248, 253, 270, 275, 412
Liddle (2004), 246, 412
Lidsey et al. (1997), 135, 247, 248, 253, 254, 256, 260, 261, 269, 412
Lidsey et al. (2000), 159, 413
Lidsey et al. (2004), 274, 413
Lie derivative, 287
line bundle, 375
Liouville equation, 344
local cross section, 373
loop corrections, 213
loop quantum gravity, 273
Lorentzian section, 213
Lovelock (1971), 16, 381, 413
Lukas et al. (1999a), 159, 160, 167, 315, 413
Lukas et al. (1999b), 159, 160, 167–169, 171, 172, 315, 413
Lukas et al. (1999c), 85, 159, 160, 167, 168, 315, 413
Lyth and Riotto (1999), 89, 119, 153, 413
Lyth et al. (2004), 127, 413
Lyth (2002a), 179, 413
Lyth (2002b), 179, 413

M-theory, 15, 84
Maartens et al. (2000), 247, 269, 413
Maartens (2004), 161, 413

Maccio et al. (2003), 211, 413
Macdonald function, 231
magnetic induction, 51
Majorana conjugation, 20
Majorana spinor, 20
Maldacena (1998), 180, 274, 413
Maldacena (2003a), 129, 132, 413
Maldacena (2003b), 180, 274, 413
manifold, 367
marginal operator, 399
Martel et al. (1998), 147, 414
Martin and Brandenberger (2002), 123, 124, 414
massive vector boson, 73
massless vector boson, 73
matter domination, 93
matter spectral index, 104
Mauro (2003), 27, 414
Maxwell–Bolzmann tail, 100
Melchiorri and Odman (2002), 148, 414
Melchiorri et al. (2003), 149, 414
Merzbacher (1998), 384, 386, 390, 414
metric, 363
metric adjoint, 366
metric connexion, 379
Mezincescu and Townsend (1985), 181, 195, 287, 414
Minces and Rivelles (2001), 195, 200, 414
minimally supersymmetric Standard Model, 119
mixed state, 27
modified gravitational action, 281
moduli fields, 191
momentum modes, 80
momentum space Feynman rules, 42
monopoles, 89, 124
Morse and Feshbach (1953), 202, 257, 350, 414
Moss (2003), 169, 170, 414
Moss (2004), 170, 414
Motl (2003), 197, 414
Moyal, 29
Mukhanov equation, 130
Mukhanov et al. (1992), 89, 129, 136, 245, 414
Mukhanov–Sasaki, 131
Mukhanov–Sasaki variable, 128
Mukhanov (1985), 128–130, 414
Mukhanov (1998), 129, 414
Mukohyama et al. (2000), 166, 262, 286, 287, 290, 315, 414

negative norm states, 72
neutron decay, 101
Neveu–Schwarz, 74
Noether theorem, 18
Nojiri and Odintsov (2000), 250, 414
Nojiri and Odintsov (2001), 276, 414
Nojiri and Odintsov (2002), 276, 414
Nojiri and Odintsov (2003), 276, 414
Nojiri et al. (2002), 250, 415
nucleosynthesis, 100
  constraints, 211
null states, 70
number density, 94

Olde Daalhuis (2001), 336, 415
Olver (1974), 336, 415
one-particle irreducible diagrams, 47
open string tachyon, 74, 191
open strings, 68
operator product expansion, 28, 137
orbifold fixed planes, 178
orthonormal basis, 364
Oscillation theorem, 347
Ovrut et al. (2000), 174, 415

Papapetrou–Majumdar metric, 176
parallel transporter, 376
Parkinson et al. (2003), 212, 415
particle horizon, 122

partition function, 36
Path integral, 352
Peacock (1999), 88, 89, 92, 113, 115, 118, 123, 124, 142, 392, 415
Peebles (1993), 90, 415
Penzias and Wilson (1965), 152, 415
Percival et al. (2001), 154, 415
Percival et al. (2002), 154, 415
Perez-Victoria (2001), 182, 276, 303, 311, 415
period, 370
Perry (1982), 189, 415
Perry (1984), 183, 416
Perry (2001), 189, 416
perturbation series, 34
Peskin and Schroeder (1995), 24, 67, 393, 416
phantom matter, 149
phase, 389
phase space, 24
photon diffusion, 109
physical phase space, 53
physical states, 59, 70
pivot wavenumber, 138
Poincaré dual, 365
Poincaré star, 365
Poisson bracket, 25
Poisson equation, 62
Polarski and Starobinsky (1996), 383, 385, 390, 416
Polchinski (1998), 24, 28, 61, 65, 66, 68–70, 75, 81, 82, 84, 168, 178, 273, 416
Polnarev (1985), 246, 416
Polyakov action, 64
power law inflation, 135
power spectrum, 104, 137
Prüfer form, 345
principal fibre bundle, 373
projection, 373
projection operator, 48
pull-back, 362
pure state, 27
pure Yang–Mills theory, 379
push-forward, 362

quantization, 283
quantum chromodynamics, 48
quantum constraints, 283
quantum cosmology, 273
quantum effective action, 46, 215
quantum effective potential, 47, 213, 215
quantum electrodynamics, 63
quantum FRW metric, 275
quantum gravity
  semi classical, 273
quasi-invariant Lagrangian, 57
quasi-particle operator, 386
quasi-particles, 386
quintessence, 149, 209
  effective action, 215
  expectation values, 210
  problems, 210

radiation domination, 93
radiative corrections, 40, 213
Ramírez and Liddle (2003), 262, 416
Ramond–Neveu–Schwarz superstring, 76
Randall and Sundrum (1999a), 160, 182, 315, 416
Randall and Sundrum (1999b), 160, 183, 197, 227, 315, 319, 416
Randall–Sundrum model, 165, 197, 227
Randall–Sundrum propagator, 227
Randall–Sunrum model, 199
Rattazzi and Zaffaroni (2001), 396, 416
Raychaudhuri equation, 91, 144, 188
Rayleigh quotient, 348
real polarization, 32
Reall (2003), 191, 337, 416

recombination, 99
reduced phase space, 50
Regge slope, 65
regularization, 353
reheating, 95
relevant operator, 399
renormalizable theory, 394
renormalization, 213, 353
renormalization group, 211, 393
  fixed point, 395
renormalized mass, 41
Ricci identity, 167
Ricci tensor, 16
Riemann $\zeta$-function, 355
Riemann curvature tensor, 379
Riesz and Sz.-Nagy (1955), 251, 259, 343, 416
Riotto (2002), 89, 126, 129, 416
Rivers (1988), 24, 46, 416
Robertson–Walker metric, 90
Rosu et al. (1999), 305, 416
Rovelli (1997), 274, 416

Saha's equation, 100
Sahni and Starobinsky (2000), 147, 416
Santos et al. (2001), 190, 416
Sanyal (2003), 276, 417
Sasaki (1986), 128, 129, 131, 417
scalar deformation, 125
scalar–tensor gravity, 210
scale invariant spectrum, 124
Schmidt (1997), 338, 417
Schoen and Yau (1979), 189, 417
Schrödinger functional, 283
Scranton et al. (2003), 210, 417
Seahra et al. (2003), 276, 277, 286, 296, 417
Seahra (2003), 277, 417
second quantization, 23
second-fundamental form, 279
sections, 213
Seery and Bassett (2004), 3, 417
Seery and Taylor (2003), 3, 315, 417
Seery (2001), 179, 417
self-adjoint operators, 344
self-energy, 40
semi-classical quantum gravity, 273
semi-classicality
  of quantum fluctuations, 383
Sen (1998), 74, 191, 417
Sen (1999), 74, 191, 417
Sen (2002a), 74, 191, 417
Sen (2002b), 74, 191, 417
Sen (2003), 74, 191, 417
shift vector, 279
Shiromizu et al. (2000), 184, 417
Shoemaker (2004), 246, 417
sigma model action, 75
signature, 364
Silk damping, 109
simplicial homology, 370
Sloan digital sky survey, 152
slow-roll approximation, 135
slow-roll parameters, 122
special holonomy, 167
spectral index, 137
Spergel et al. (2003), 100, 137, 210, 245, 418
spin connexion, 380
Spradlin et al. (2001), 201, 396, 418
squeezed states, 386
squeezing angle, 389
squeezing parameter, 389
Starobinsky (1985), 246, 418
Steinhardt and Turok (2001), 175, 315, 418
Steinhardt and Turok (2002), 175, 418
Stephani et al. (2003), 313, 418
Sternheimer (1998), 29, 418
Stewart and Lyth (1993), 135, 256, 260, 418

Stewart (1991), 18, 190, 358, 418
Stokes' theorem, 280, 363
string compactifications, 75
string Fadeev–Popov operator, 66
string theory, 64
  matter theory, 65
Strominger (2001), 201, 396, 418
strong energy condition, 190
structure constants, 378
structure group, 373
Stueckelberg and Peterman (1952), 393, 418
Sturm–Liouville, 250
  density, 344
  Hilbert space, 345
  inner product, 345
  measure, 345
  orthonormality of eigenfunctions, 345
  reality of weights, 345
  regular, 345
  singular, 345
  weight, 344
Sturm–Liouville expansion of functional measure, 353
Sturm–Liouville operator, 251
Sturm–Liouville problem, 344
Sturm–Liouville transform, 352
super-Hamiltonian, 282
supersymmetry, 76, 167
surface of last scattering, 112
Susskind et al. (1993), 289, 303, 418
Susskind (1995), 124, 418
Susskind (2003), 151, 207, 418
Susskind (2004), 151, 418
Symanzik (1970), 394, 418
symplectic form, 25
symplectic manifold, 25
symplectic product, 194
synchronous gauge, 105
T-duality, 80
tachyonic fields, 191
tachyonic vector, 73
tangent bundle, 375
tangent mapping, 362
tangent space, 360
tangent vector, 360
Temme (2001), 336, 418
tensor product, 361
Thiemann (2003), 274, 419
three-body interactions, 102
three-geometry, 279
time evolution operator, 31
Tocchini-Velentini and Amendola (2002), 211, 419
Tolley et al. (2004), 179, 419
topological, 89
torsion, 379
total manifold, 373
Townsend (2001), 208, 419
transplanckian problem, 123
tree-level potential, 213
tunelling, 275
Turok et al. (2004), 177, 316, 419
Type I string theory, 83
Type II string theory, 83, 178
Tyutin (1975), 56, 419
ultra-violet completion, 158
ultra-violet divergence, 43
uncertainty relation, 383
uniform curvature hypersurfaces, 126
uniform density hypersurfaces, 126
unitarity, 63
unrenormalizable theory, 394
vacuum parameters, 139
vacuum persistance amplitude, 34
vacuum–vacuum amplitude, 34

variance, 137
vector bundle, 375
vector potential, 51
vector space
  dual, 358
Verlinde compactification, 179
Verlinde (2000), 182, 252, 311, 319, 419
Verlinde (2003), 197, 419
Vernizzi (2004), 127, 419
vielbein, 360
Vilenkin and Shellard (), 89, 419
Virasoro algebra, 28
Virasoro generators, 69, 79
virtual particles, 214
Visser (1996), 21, 22, 157, 419

Wainwright and Ellis (1997), 117, 419
Wald (1983), 89, 144, 419
Wald (1984), 179, 419
Wald (1994), 24, 140, 194, 195, 419
Wands et al. (2000), 125–128, 179, 250, 419
Ward identity, 52
warped compactifications, 197
Watson (1918), 336, 420
wedge product, 359
Weinberg (1972), 16, 88–92, 103, 117, 166, 183, 420
Weinberg (1975), 98, 420
Weinberg (1987), 147, 420
Weinberg (1994), 18, 23, 24, 46, 58, 59, 67, 125, 137, 167, 170, 213, 232, 234, 237, 249, 338, 353, 380, 381, 393, 395, 420
Weinstein (1994), 29, 420
Weyl spinors, 20
Weyl symbol, 28
Weyl tensor, 184
Weyl transformation, 65, 381
Wheeler–de Witt equation, 283
White and Hu (1997), 113, 115, 420
Wick rotation, 43, 241
Wick's theorem, 354
Wiener–Khinchin, 104
Wilkinson Microwave Anisotropy Probe, 152
Wiltshire (1996), 301, 420
winding modes, 80
winding number, 78
Witten (1981), 169, 189, 190, 420
Witten (1996), 85, 175, 182, 420
Witten (1998), 195, 200, 420
Witten (1999a), 182, 200, 420
Witten (1999b), 24, 420
Witten (1999c), 174, 420
Witten (2003), 44, 420
Wong and Guo (1989), 336, 420
world sheet, 64
Wronskian, 230

Yang–Mills, 378
Yang–Mills theory, 15, 47
Yukawa force, 210

Zaldarriaga and Seljak (1997), 246, 420
Zel'dovich (1966), 98, 421
zero modes, 75